Promotors	Prof. dr. ir. Niko Verhoest
	Prof. dr. ir. Valentijn Pauwels
	Laboratory for Hydrology and Water Management,
	Faculty of Bioscience Engineering, Ghent University
Dean	Prof. dr. ir. Herman Van Langenhove
Rector	Prof. Paul Van Cauwenberge

Gabriëlle De Lannoy

Assimilation of Soil Moisture Observations into a Spatially Distributed Hydrologic Model

Thesis submitted in fulfillment of the requirements for the degree of Doctor (PhD) in Applied Biological Sciences

Dutch translation of the title:
Assimilatie van bodemvochtwaarnemingen in een ruimtelijk gedistribueerd hydrologisch model

Cover:
Front: Soil moisture modeling with assimilation of observations at probe BL2 in the corn field of the Optimizing Production Inputs for Economic and Environmental Enhancement (OPE^3) project, Beltsville, Maryland, USA.
Back: Photograph taken in field A of the OPE^3 field.

Reference:
De Lannoy, G. J. M. (2006). Assimilation of Soil Moisture Observations into a Spatially Distributed Hydrologic Model. PhD.-dissertation, Ghent University, 548p.

ISBN-number: 90-5989-141-4
ISBN-number: 9789059891418

Numquam aliud natura, aliud sapientia dicit

Acknowledgement

This dissertation was only possible through the help, guidance and goodwill of a number of people, who I want to thank sincerely.

To start with, I would like to thank all members of the PhD-committee, prof. dr. ir. Niko Verhoest, prof. dr. ir. Valentijn Pauwels, prof. dr. Raoul Lemeur, prof. dr. ir. Marc Van Meirvenne, prof. dr. Paul Houser, prof. dr. ir. Robin De Keyser, prof. dr. ir. Ann van Griensven and prof. em. dr. ir. François De Troch, for their time to deal with this dissertation and for providing me some good feedback. I thank my advisors, Niko and Valentijn, for their multiple and patient reading of this text and for leaving me so much freedom in my research. I also thank prof. De Troch and Niko for giving me the opportunity to do PhD-research at the LHWM and Marc for his contribution to the spatial analysis of the field observations in my study. I am most indebted to prof. dr. Paul Houser. He basically offered all the ingredients for this work: the access to the data, the suggestion for the model, good ideas, encouragement and assistance in paper writing. With a few bright remarks on my trials during a few meetings, Paul could inspire me for long afterwards. Together with my advisors, he arranged a visit to the HSB of NASA/GSFC in 2002. This was an unforgettable experience, thanks to many kind people: many thanks to all of them! It was during this stay that my PhD-research was put on track with the help of, amongst others, prof. dr. ir. Jeffrey Walker, who kindly proposed an initial data set and a model, and dr. ir. Rolf Reichle, who provided insightful research directions. I thank Rolf for helping intensively during the last weeks of my research to sort out and clarify results for a publication.

This research completely builds on to data kindly provided by the USDA-ARS Hydrology and Remote Sensing Laboratory (dr. Timothy Gish and dr. William Kustas) and the Farm Operations Branch (George Meyers). Special thanks go to Tim for sharing his insights in the OPE^3 field and to Andrew Russ for helping me in organizing the dataset.

Financial support was provided through research grant BOF 011D2901 from the Special Research Fund (BOF) of the Ghent University. This research was also partially funded through STEREO grant SR/00/01 of the Belgian Science Policy (BELSPO).

I would like to thank my cousin ir. Michaël Steenbergen for his kind help on the spectral analysis for the up-scaling of the observations, which in the end was excluded from this dissertation, but entirely used for a publication. My current (Jan, Marcella, Dave) and former (Annelies, Christophe, Inge, Moira, Valerie, Vanessa) colleagues helped very well to detach my thoughts from an abstract world and Rudi guaranteed a stable computer environment.

Fortunately, also a PhD-student gets a life outside the office. I am very grateful for all good moments with friends during the weekly orchestra, flute and gymnastics lessons. Also, I will not forget the nice scout camps and the good education by my colleagues of the Red Cross.

In closing, I owe a lot of gratitude to my parents, my brothers, Lorenzo and Jonas, and sister Yacintha. Their questionless support and devotion helped me very much to persevere. Dear brothers, thank you for many things, from the cover pictures to washing my car and joining me for a great mountain trip. 'Little' sister, special thanks for being so amusing and for keeping my Latin and maths so fresh! A sincere word of thanks goes to my mother, who did about everything to make my life and work easier, when I was trying to break time records.

Ghent, November, 2006.
Age quod agis, cura nihil aliud nisi ut valeas.

Abstract

The assimilation of soil moisture observations for parameter, state and forecast bias estimation in a land surface model was explored. The test system was the agricultural field on which the Optimizing Production Inputs for Economic and Environmental Enhancement (OPE^3) project is conducted. The analysis of 1 year soil moisture time series measured for 36 profiles resulted in a number of spatio-temporal soil moisture characteristics, which could be useful in a further stage of system modeling and assimilation. The Community Land Model (CLM2.0) was used to represent the system as a collection of independent one-dimensional land columns. The optimization of the parameters and initial states was performed in a multi-objective framework by weak constraint variational assimilation of soil moisture profile observations during one month.

To assess the uncertainty in the a priori estimated soil moisture states, statistics of ensemble model simulations were studied after combined or separate perturbation of initial states, parameters and/or forcings. Through perturbation of the initial states only, no realistic forecast uncertainty could be simulated and model error was the major source of uncertainty in the forecasted states. Ensemble perturbation sometimes introduced bias in the mean a priori state estimates.

The ensemble model results were combined with observations to update the state estimate along the model simulations by an ensemble Kalman filter. Because the model forecasts showed the presence of persisting errors, i.e. bias, the filtering algorithm for state estimation was extended with several filter variants for bias estimation. The combined state and bias estimation improved the precision of the a posteriori soil moisture state estimates, but sometimes negatively influenced the depending output fluxes and water balance. Depending on the kind of model bias, a different estimation scheme could be proposed. The algorithms could be further enhanced to overcome a number of unrealistic assumptions in the bias estimation procedure.

Contents

List of Symbols

Symbol	Description
$\bar{\cdot}$	spatial or ensemble average, as indicated in the text
$\overline{\cdot}^{\mathrm{x}}$	spatial average of a variable
$\bar{\cdot}_i$	ensemble average of a variable, at time step i
$E[\cdot]$	expected value operator
$< \cdot >$	temporal average
$tr(\cdot)$	trace of a matrix, sum of the diagonal elements
$\mathbf{1}$	matrix with all elements equal to 1
$\mathbf{I}$	unity matrix (elements on diagonal are 1)
$\gamma(h)$	variogram as function of the lag h
$\tan\beta$	local surface topographic slope
λ	topographic index
μ	mean value
σ	standard deviation
$\mathbf{B}_i$	linear operator, matrix relating input to the state, at time step i
$\mathbf{b}_i^a$	analysis bias vector, at time step i
$\mathbf{b}_i^f$	forecast bias vector, at time step i
$\hat{\mathbf{b}}_i^f$	a posteriori forecast bias estimate vector, at time step i
$\hat{\mathbf{b}}_i^{f,-}$	a priori forecast bias estimate vector, at time step i
$\hat{\mathbf{b}}_{j,i}^{f,-}$	ensemble member j of the a priori bias estimate vector, at time step i
$\hat{\mathbf{b}}_{l,i}^f$	a posteriori forecast bias estimate vector in land column l, at time step i
$\hat{\mathbf{b}}_{l,i}^{f,-}$	a priori bias estimate vector in land column l, at time step i
$\bar{\hat{\mathbf{b}}}_i^{f,-}$	ensemble mean a priori bias estimate vector, at time step i ($\equiv \hat{\mathbf{b}}_i^{f,-}$)
$\mathbf{b}_i^o$	observation bias vector, at time step i
$\mathbf{e}_i$	a posteriori state estimation error vector, at time step i
$\mathbf{e}_i^-$	a priori state estimation error vector, at time step i
$\tilde{\mathbf{e}}_i$	zero mean a posteriori state estimation error vector, calculated based on biased information, at time step i

Symbol	Description
$\tilde{\mathbf{e}}_i^-$	zero mean a priori state estimation error vector, calculated based on biased information, at time step i
$\tilde{\tilde{\mathbf{e}}}_i$	non-zero mean a posteriori state estimation error vector, due to a biased a posteriori estimate, at time step i
$\tilde{\tilde{\mathbf{e}}}_i^-$	non-zero mean a priori state estimation error vector, due to a biased a priori estimate, at time step i
$\mathbf{e}_i^{-,p}$	predictability error vector, due to analysis error, at time step i
$\mathbf{F}_{i+1,i}$	linear operator, transition matrix propagating the state at time step i to $i+1$
$\mathbf{f}_{i+1,i}(\mathbf{x},\dots)$	discrete non-linear system function for the transition of the state at time step i to $i+1$
$\boldsymbol{\mathcal{F}}''_{i+1,i}$	Hessian matrix of the system function $\mathbf{f}_{i+1,i}()$
$\mathbf{G}_i$	linear operator relating the model noise to the state, at time step i
$\mathbf{g}_{i,i-1}$	function propagating the bias vector from time step $i-1$ to i
$\mathbf{H}_i$	linear operator, matrix transforming the state to the output/observation, at time step i
$\mathbf{h}_i()$	non-linear function relating the state to the output, at time step i
$\mathbf{K}(t)$	gain matrix function used with continuous systems (e.g., nudging)
$\mathbf{K}_i$	weighting or blending matrix, at time step i (mostly Kalman gain)
$\tilde{\mathbf{K}}_i$	Kalman gain matrix based on a priori error covariances for biased variables, at time step i
$\mathbf{K}_{l,i}$	Kalman gain matrix to update the state estimate for an individual land columns l, at time step i
$\mathbf{K}_i^k$	weighting matrix at the kth iteration, at time step i (successive corrections)
$\mathbf{K}_{b^f,i}$	Kalman gain matrix to update the bias estimate, at time step i
$\mathbf{K}_{b^f,l,i}$	Kalman gain matrix to update the bias estimate in land column l, at time step i
$\mathbf{k}_i^k$	vector of weights, at time step i, at the kth iteration (successive correction)
$\mathbf{k}(\mathrm{x},\mathrm{y},\mathrm{z},t)$	space and time dependent weighting vector (nudging)
$\mathbf{P}_i$	a posteriori state estimation error covariance matrix, at time step i
$\mathbf{P}_i^-$	a priori state estimation error covariance matrix, at time step i
$\mathbf{P}_i^{-,p}$	predictability error covariance matrix, at time step i
$\mathbf{P}_\mathbf{C}$	time-invariant correlation matrix (optimal interpolation)
$\mathbf{P}_{\mathbf{S}i}$	diagonal matrix of forecast error standard deviations, at time step i (optimal interpolation)

Symbol	Description
$\tilde{\mathbf{P}}_i$	a posteriori bias-blind state estimation error covariance matrix, at time step i
$\tilde{\mathbf{P}}_i^-$	a priori bias-blind state estimation error covariance matrix, at time step i
$\tilde{\tilde{\mathbf{P}}}_i$	a posteriori bias-blind state estimation error correlation matrix, at time step i
$\tilde{\tilde{\mathbf{P}}}_i^-$	a priori bias-blind state estimation error correlation matrix, at time step i
$\mathbf{P}_{bf,i}$	a posteriori bias estimation error covariance matrix, at time step i
$\mathbf{P}_{bf,i}^-$	a priori bias estimation error covariance matrix, at time step i
$\mathbf{P}_{members,i}$	covariance matrix between ensemble members, at time step i
$\mathbf{P}_{states,i}$	covariance matrix between state variables, at time step i
$\mathbf{Q}_i$	model noise covariance matrix, at time step i
$\mathbf{R}_i$	observation error covariance matrix, at time step i
$\tilde{\mathbf{R}}_i$	observation error covariance matrix for biased observations, at time step i
$\tilde{\tilde{\mathbf{R}}}_i$	observation error correlation matrix for biased observations, at time step i
$\mathbf{r}_i$	residual (a posteriori) vector, at time step i
$\mathbf{r}_i^-$	innovation or a priori residual vector, at time step i
$\mathbf{r}_{j,i}^-$	ensemble member j of the innovation vector, at time step i
$\bar{\mathbf{r}}_i^-$	ensemble mean innovation vector, at time step i
$\mathbf{u}_i^*$	input vector at a working point for linearization, at time step i
$\mathbf{u}_i$	input vector, at time step i
$\mathbf{V}_i$	Jacobian matrix, containing partial derivatives $\frac{\partial \mathrm{h}_c}{\partial v_s}$, at time step i
$\mathbf{v}_{j,i}$	ensemble member j of the observation noise vector, at time step i
$\mathbf{v}_i$	observation noise vector, at time step i
$\tilde{\tilde{\mathbf{v}}}_i$	observation error vector for biased observations, at time step i
$\mathbf{W}_i$	Jacobian matrix, containing partial derivatives $\frac{\partial \mathrm{f}_a}{\partial w_r}$, at time step i
$\mathbf{w}_{j,i}$	ensemble member j of the model noise vector, at time step i
$\mathbf{w}_i$	model noise vector, at time step i
$\hat{\mathbf{X}}_i^-$	matrix with columns of ensemble members of the a priori estimated state vector, at time step i
$\hat{\mathbf{X}}_i$	matrix with columns of ensemble members of the a posteriori estimated state vector, at time step i
$\hat{\mathbf{X}}_i'$	matrix with columns of ensemble perturbations around the state vector, data matrix, at time step i

Symbol	Description
$\mathbf{x}_i$	true state, or control vector only in case of variational assimilation, at time step i
$\mathbf{x}_i^*$	state vector at a working point for linearization, at time step i
$\hat{\mathbf{x}}_i$	a posteriori state estimate vector, at time step i
$\hat{\mathbf{x}}_i^-$	a priori state estimate vector, at time step i
$\bar{\hat{\mathbf{x}}}_i$	ensemble mean a posteriori estimated state, at time step i ($\equiv \hat{\mathbf{x}}_i$)
$\bar{\hat{\mathbf{x}}}_i^-$	ensemble mean a priori estimated state, at time step i ($\equiv \hat{\mathbf{x}}_i^-$)
$\hat{\mathbf{x}}_i^{-,-}$	a priori state estimate vector, corrected for a priori estimated bias vector, at time step i
$\mathbf{x}_{l,i}$	partial state vector for land column or profile l, at time step i
$\hat{\mathbf{x}}_{j,i}$	ensemble member j of the a posteriori estimated state vector, at time step i
$\hat{\mathbf{x}}_{j,i}^-$	ensemble member j of the a priori estimated state vector, at time step i
$\hat{\mathbf{x}}_{l,j,i}$	ensemble member j of the a posteriori state estimate vector for the individual land columns l, at time step i
$\hat{\mathbf{x}}_{l,j,i}^-$	ensemble member j of the a priori state vector for the individual land columns l, at time step i
$\hat{\tilde{\mathbf{x}}}_i$	biased a posteriori state estimate vector, at time step i
$\hat{\tilde{\mathbf{x}}}_i^-$	biased a priori state estimate vector, at time step i
$\hat{\tilde{\mathbf{x}}}'_i$	partially bias-blind a posteriori state estimate vector, at time step i (EnBKF_2/2^+)
$\hat{\tilde{\mathbf{x}}}'^-_i$	partially bias-blind a priori state estimate vector, at time step i (EnBKF_2/2^+)
$\hat{\tilde{\mathbf{x}}}_{j,i}$	ensemble member j of the a posteriori biased state estimate vector, at time step i
$\hat{\tilde{\mathbf{x}}}_{j,i}^-$	ensemble member j of the a priori biased state estimate vector, at time step i
$\tilde{\mathbf{x}}_i$	true state vector, with inclusion of mean forecast bias (auxiliary variable), at time step i
$\bar{\hat{\tilde{\mathbf{x}}}}_i^-$	ensemble mean a priori biased state estimate, at time step i ($\equiv \hat{\tilde{\mathbf{x}}}_i^-$)
$\boldsymbol{x}_i$	dummy variable in the probability density function, at time step i
$\mathbf{Y}_i$	matrix with columns of perturbed realizations of a measurement vector, at time step i
$\mathbf{y}_{j,i}$	ensemble member j of the observation vector, at time step i
$\mathbf{y}_i$	output or observation vector, at time step i
$\tilde{\mathbf{y}}_i$	biased observation vector, at time step i
A	area of the upslope watershed per unit contour length
a	range of a variogram
$C(0)$	nugget of a variogram
$C(h)$	spatial covariance function, as function of the lag h

Symbol	Description
C_l	sill of a variogram
CV	coefficient of variation
$c_{j,i}^k$	Cressman weight for the jth observation point at the kth iteration, at time step i (successive correction)
$c_{\mathrm{x,y,z},t}$	space and time dependent weighting function (nudging)
d_j	distance between the jth observation point and the grid point for the analysis (successive correction)
$d_{j,i}$	(relative/absolute) difference between soil moisture at point j and spatially averaged soil moisture, at time step i (time stable sites)
e_a	actual vapor pressure [Pa]
e_s	saturated vapor pressure [Pa]
$ensk_i$	ensemble skill, at time step i
$ensp_i$	ensemble spread, variance or dispersion, at time step i
$f_{\mathbf{x}_i}(\boldsymbol{x})$	probability density function of $\mathbf{x}_i$, with $\boldsymbol{x}$ a realization of $\mathbf{x}$
$f(x,t)$	continuous non-linear function propagating a scalar state variable x in time t (e.g., nudging)
$\mathrm{f}_{a,i+1,i}(x,..)$	similar as $\mathbf{f}_{i+1,i}(\mathbf{x},\dots)$ but for a scalar $x_{a,i}$, instead of for a vector $\mathbf{x}_i$
g_x	nudging coefficient (nudging)
h	lag or distance, used for spatial correlation functions (incl. variograms)
$\mathrm{h}_{c,i}(x,\dots)$	similar as $\mathbf{h}_i(\mathbf{x})$ but for a scalar $y_{c,i}$, instead of for a vector $\mathbf{y}_i$
$\mathcal{J}$	cost functional, penalty (variational assimilation)
$\mathcal{J}_o$	observation penalty (variational assimilation)
$\mathcal{J}_b$	background penalty (variational assimilation)
$k_{j,i}^k$	element of $\mathbf{k}_i^k$ for observation j (successive correction)
$k(\mathrm{x},\mathrm{y},\mathrm{z},t)$	element of $\mathbf{k}(\mathrm{x},\mathrm{y},\mathrm{z},t)$ (nudging)
$kurt_{q,i}$	multivariate kurtosis, based on a rank q pseudo-inverse of the ensemble covariance matrix, at time step i
L_E	correlation length, where a simple exponental correlation function reaches 1/e (=0.368)
L_S	correlation length, where a compactly supported second order autoregressive correlation function reaches 2/e (=0.736)
M	median value
m	mean value
m	dimension of $\mathbf{y}_i$
mse_i	mean squared error of all the ensemble forecasts, at time step i
NL	number of layers in a profile
N	number of ensemble members
$n(h)$	number of pairs of data separated by lag h, used to calculate the variograms
n	dimension of $\mathbf{x}_i$
p	pressure [Pa]
p	dimension of $\mathbf{u}_i$

Symbol	**Description**
q	estimate of the ratio of the observation error to the background error covariance (successive corrections)
$q_{\mathrm{x,y,z}}$	quality factor (nudging)
R	correlation, calibration/validation measure
RH	relative humidity [-]
R_i^k	radius of influence at the estimation time step i, mostly shrinking for successive iterations k (successive correction)
$r(h)$	spatial correlation function
$r_{z_k z_l}$	zero-time lag correlation between time series of variable z_k and z_l (cross-correlation)
r	correlation coefficient (regression)
r^2	determination coefficient (regression)
SH	specific humidity [-]
$skew_{q,i}$	multivariate skewess, based on a rank q pseudo-inverse of the ensemble covariance matrix, at time step i
$stdv$	standard deviation
T	temperature [℃]
$\mathrm{T}_{(\mathrm{L})}$	number of time steps (in layer L)
$u_{b,i}$	salar element of $\mathbf{u}_i$, at time step i
$\mathrm{x}, \mathrm{y}, \mathrm{z}$	spatial coordinates
$\hat{x}_i^k$	scalar element of a posteriori state estimate vector at the kth iteration, at time step i (successive correction)
$x_{a,i}$	salar element of $\mathbf{x}_i$, at time step i
$y_{c,i}$	salar element of $\mathbf{y}_i$, at time step i

Glossary

Notation	Description	
$BIAS$	(absolute) mean difference (as objective function)	116
N_S	Nash-Suttcliffe criterium (as objective function)	116
R	Time series correlation coefficient (as objective function)	116
$RMSE$	Root Mean Square Error (as objective function)	116
stdv	standard deviation	46
1D	One-dimensional	2
3D	Three-dimensional	2
3D-Var	Three-dimensional Variational assimilation	29
4D-Var	Four-dimensional Variational assimilation	30
ARMA	Auto Regressive Moving Average	295
ARMAX	Autoregressive Moving Average with eXogenous input	119
ARS	Agricultural Research Service	38
BARC	Beltsville Agricultural Research Center	38
BATS	Biosphere-Atmosphere Transfer Scheme	114
BLUE	Best Linear Unbiased Estimator	18
CAM	Community Atmosphere Model	124
CASMM	Catchment Averaged Soil Moisture Monitoring	92
CCSM	Community Climate System Model	124
cdf	cumulative distribution function	6
CLM	Community Land Model	48
CSI	Campbell Scientific Inc.	46
DA depth	Data Assimilation depth	309
DA event	Data Assimilation event	310
DA frequency	Data Assimilation frequency	309
DEM	Digital Elevation Model	39
DMIP	Distributed Modeling Intercomparison Project	109
DOQ	Digital Orthophoto Quadrangle	47
DOY	Day Of the Year	48

Notation	Description	
ECMWF	European Centre for Medium-Range Weather Forecasts	216
EKF	Extended Kalman filter	23
ELDAS	European Land Data Assimilation System	294
EnBKF_0	bias estimation only, without state update	363
EnBKF_1	separate state and bias estimation without bias feedback	363
EnBKF_2/2^+	separate state and bias estimation with partial bias feedback on the analysis through bias-corrected innovations	363
EnBKFf_a	separate state and bias estimation with complete bias feedback on the analyses and forecasts	363
EnBKFf_b/b^+	separate state and bias estimation with complete bias feedback on the analyses	364
EnKF	Ensemble Kalman filter	25
Ens Mean	Ensemble Mean integration without filtering	310
EPA	United States Environmental Protection Agency	40
EPS	Ensemble Prediction System	215
GCM	General Circulation Model	108
GEOS DAS	Goddard Earth Observing System Data Assimilation System	339
GIS	Geographic Information System	43
GLDAS	Global Land Data Assimilation System	294
GLUE	Generalized Likelihood Uncertainty Estimation	118
GPR	Ground Penetrating Radar	55
GPS	Global Positioning System	39
GSFC	Goddard Space Flight Center	38
H	sensor at location with High infiltration	55
HEPEX	Hydrological Ensemble Prediction Experiment	219
HMLE	Heteroscedastic Maximum Likelihood Estimator	116
HSB	Hydrological Sciences Branch	38
IAP94	LSM developed at the Institute of Atmospheric Physics	124
IGBP	International Geosphere-Biosphere Programme	170
L	sensor at location with Low infiltration	55
LDAS	Land Data Assimilation System	294
LSM	Land Surface Model	108
M	sensor at location with Moderate infiltration	55
MC	Monte Carlo	117
MOCOM	Multi-Objective Complex Evolution	117
MODIS	Moderate Resolution Imaging Spectroradiometer	293
MOGSA	Multi-Objective Generalized Sensitivity Analysis	114

Notation	Description	
MOPSO	Multi-Objective Particle Swarm Optimization	117
MOSCEM	Multi-Objective Shuffled Complex Evolution Metropolis	117
MSE	Mean Squared Error	116
NAD83	North American Datum of 1983	42
NASA	National Aeronautics and Space Administration	38
NCAR	National Center for Atmospheric Research	124
NCEP	National Center for Environmental Prediction	216
NDVI	Normalized Difference Vegetation Index	153
nir	near infrared radiation	152
NLDAS	North American Land Data Assimilation System	294
NOAA	National Oceanic and Atmospheric Administration	109
NRCS	Natural Resources Conservation Service	39
NWP	Numerical Weather Prediction	214
NWS	National Weather Service (NOAA)	109
OBA	Old Beltsville Airport	48
OI	Optimal Interpolation	18
OPE^3	Optimizing Production Inputs for Economic and Environmental Enhancement	38
OSSE	Observation System Simulation Experiment	216
PDE	Partial Differential Equations	107
pdf	probability density function	6
pft	plant functional type	125
PILPS	Project for Intercomparison of Land surface Parameterization Schemes	109
PPS	Probabilistic Prediction System	224
PPU	Potential Predictive Utility	228
PSAS	Physical Space Assimilation System	33
PSO	Particle Swam Optimization	113
ROC	Relative Operating Characteristic	223
RTM	Radiative Transfer Model	283
SΔE	Stochastic Difference Equation	11
SAC-SMA	Sacramento Soil Moisture Accounting model	116
SCA	Snow Covered Area	293
SCAN	Soil Climate Analysis Network	39
SCE	Shuffled Complex Evolution	113
SCM	Successive Corrections Method	15
SDE	Stochastic Differential Equation	11
SHE	Système Hydrologique Européen	114
SHETRAN	extended version of SHE	114

Notation	**Description**	
SM	Soil Moisture	61
SSE	system simulation experiment	217
SVAT	Soil-Vegetation-Atmosphere Transfer	108
SVD	Singular Vector Decomposition	308
SWE	Snow Water Equivalent	293
TDR	Time Domain Reflectometer	282
TOPLATS	TOPMODEL based Land Atmosphere Transfer Scheme	117
USDA	United States Department of Agriculture	38
USGS	United States Geological Survey	39
vis	visible radiation	152
WWW	World Wide Web	39

Chapter 1

Introduction

Soil moisture has a large impact on diverse physical and chemical earth processes. Estimation of land surface variables and soil moisture in particular is of major importance to understand the partitioning of the heat and water fluxes in land systems. Several authors have stressed the importance of soil moisture at the global scale for weather [Zhang and Frederiksen, 2003; Koster *et al.*, 2004] and climate predictions [Dirmeyer, 2000]. At the local scale, soil moisture has a large impact on runoff and flooding, vegetation and agricultural yield, biogeochemical processes, soil stability, etc.. A good estimate of the land surface state is needed for good predictions, to understand the processes within the system, and to allow good management of controlled land systems, such as agricultural fields.

Soil moisture observations provide the most direct source of information to estimate the soil moisture state, but it is impossible to take observations that cover the different space and time scales which are spanned by the different processes researchers are interested in. Land surface models offer a complementary source of information. Through initialization and calibration, the models are adapted to represent the observed characteristics of the system. During model calibration, information from observations is basically assimilated into the model structure to obtain a best estimate for parameters. Once the model is properly calibrated and the initial conditions are well defined, the model can be used for predictions. Because model forecasts are never perfect, it is useful to combine observations with the a priori forecasted state estimates through filtering, when they become available. This should result in an improved a posteriori state estimate.

Through a weighted combination of model predictions and observations, the state can be better estimated than by either source of information individually. This merging of information to obtain a physically consistent estimate of spatially distributed environmental variables, which describe a system's state, is referred to as state estimation or state updating. In hydrology, the term 'data assimilation' has mostly been reserved for state estimation, based on McLaughlin [1995] who ar-

gued that data assimilation differed from calibration. However, data assimilation was defined more broadly by the World Meteorological Organization [WMO, 1992], and was meant for each type of assimilation of observations into model physics, including parameter estimation through calibration. In this work, the latter general definition is used and it is clearly indicated whether state or parameter estimation is discussed. Data assimilation is used in many different disciplines and aims at including real world observations in computer models.

This work studies the possibilities of and requirements for observation assimilation in hydrologic modeling to optimally estimate soil moisture in a small agricultural system. The total system is a three-dimensional (3D) land surface system, which is subdivided into a number of independent single column systems, each partitioned in a number of vertical soil layers. The state of each soil layer consists of soil moisture and temperature, but only soil moisture profile observations are available for assimilation and validation. The one-dimensional (1D) system representation, which is generally used for large regional scale modeling, allows for a computationally feasible system identification through variational assimilation, however at the expense of inevitable remaining colored model error due to the simplified process representation. Furthermore, a good understanding of the impact of different studied assimilation algorithms along the model runs benefits from a simple 1D tracking of state and update interactions only in the vertical soil columns. Furthermore, through the repetition of the methodologies for several soil columns in space, it is possible to draw general conclusions. Nevertheless, all 1D soil columns are treated simultaneously so that the total state vector consists of a number of 1D state vectors in space, but with model optimization for parallel computation. This enables us to directly extend this research to increasingly complex assimilation schemes, in which the lack of horizontal process representation is met by a sophisticated definition of the error statistics in the filtering algorithm.

The structure of this dissertation is as follows. In **chapter 2**, the general formulation of a system and its state is introduced and the spectrum of most used assimilation methods and their mathematical formulations are reviewed. Natural systems can be represented by mathematical equations, which capture most of the governing physical laws. Measurements taken in the field are basically signals, which provide information on the system. In this study, the system is an agricultural field and the main available observations are meteorological forcings (input) and soil moisture (internal state and output).

After this general chapter, the first part is devoted to the description of the observations and the land surface model which are used to characterize a small agricultural corn field, on which the Optimizing Production Inputs for Economic and Environmental Enhancement (OPE^3) project is conducted. This field is situated near Washington D.C., USA. A four-dimensional dataset of soil moisture is studied for spatial and temporal characteristics in **chapter 3**. The Community Land Model v.2.0 (CLM2.0) is chosen to represent the system and to simulate the processes for the OPE^3 field. **Chapter 4** provides some background information on system identification and includes a description of the model structure of CLM2.0.

In **chapter 5** the combined initial state and parameter estimation for CLM2.0 is discussed. The latter is basically performed in a simple weak constraint variational assimilation scheme. In this chapter, the simulated soil moisture patterns in space and time are validated through comparison with the observed spatio-temporal characteristics analyzed in **chapter 3**. A good characterization of the land surface system by the model is a prerequisite for successful state estimation through filtering along the model run.

Because deterministic forecasts do not offer an indication of the model uncertainty, ensemble simulations are generated by different types of perturbations in **chapter 6**. The ensembles are statistically interpreted and verification tools are adopted from meteorological research for weather forecasts.

The study of the ensembles and the model uncertainty provides a basis for the second part, which focuses on state and bias estimation through ensemble Kalman filtering. In **chapter 7**, mainly hydrological state estimation studies are reviewed. In **chapter 8**, a parallel ensemble Kalman filter is applied to estimate soil moisture profiles in the OPE^3 field with the CLM2.0. The influence of the assimilation frequency and depth are investigated and the assumptions underlying the procedure of Kalman filtering are studied. Based on the findings in this chapter, the need for bias estimation and correct characterization of error statistics is discussed in **chapter 9**. Several procedures for model bias correction are proposed and included in the CLM2.0. Different assimilation scenarios are studied in **chapter 10**.

In **chapter 11** the conclusions from this work are discussed and some directions for future research are given. Literature reviews are grouped by the different research items in the different chapters. Only in **chapter 2**, general topics on systems and signals are discussed for good understanding of all following chapters. The major contributions of this research can be subdivided in the following categories.

- Observation analysis: (i) the small area study of the spatio-temporal characteristics of soil moisture for an exceptionally detailed data set; (ii) the indications for use of derived statistics (and up-scaling techniques) for assimilation and validation.

- Data assimilation: (i) the grouping and interpretation of numerous assimilation techniques using a consistent notation; (ii) the step over the illogical divide built by hydrologists between calibration and data assimilation.

- System identification: (i) from the calibration point of view: extension of a multi-objective calibration algorithm for parameter estimation by an optimal initial state estimation; (ii) from the data assimilation point of view: extension of the weak constraint variational assimilation with the classical single objective function to assimilation with multiple objective functions.

- Ensembles analysis: the transfer of verification tools from meteorology to hydrology to judge the quality of soil moisture ensembles.

- State estimation: the illustration of general shortcomings to the conditions for optimal filtering with a Kalman filter in hydrologic data assimilation studies.
- State and bias estimation: (i) the development and application of a number of algorithms to simultaneously estimate the state and persisting error with an ensemble Kalman filter; (ii) the evaluation of the combined filter performances for state and bias estimation; (ii) the study of the effect of soil moisture bias correction on depending variables.

Some sections in this dissertation are based on parts of the following publications:

1. De Lannoy, G.J.M., Verhoest, N.E.C., Houser, P.R., Gish, T., Van Meirvenne, M. (2006). Spatial and temporal characteristics of soil moisture in an intensively monitored agricultural field (OPE^3). Journal of Hydrology, 331(3-4), 719-730, doi:10.1016/j.jhydrol.2006.06.016.
2. De Lannoy, G.J.M., Houser, P.R., Pauwels, V.R.N., Verhoest, N.E.C. (2006). Assessment of model uncertainty for soil moisture through ensemble verification. Journal of Geophysical Research, 111, D10101.1-18, doi:10.1029/2005 JD006367.
3. De Lannoy, G.J.M., Houser, P.R., Pauwels, V.R.N., Verhoest, N.E.C. (2006). State and bias estimation for soil moisture profiles by an ensemble Kalman filter. Water Resources Research, conditionally accepted.
4. De Lannoy, G.J.M., Reichle, R.H., Houser, P.R., Pauwels, V.R.N., Verhoest, N.E.C. (2006). Ensemble Kalman filtering of soil moisture observations with model bias correction. Water Resources Research, under review.
5. De Lannoy, G.J.M., Houser, P.R., Verhoest, N.E.C., Pauwels, V.R.N., Gish, T. (2006). Representativeness of point soil moisture observations and upscaling. Journal of Hydrology, under review.
6. De Lannoy, G.J.M., Pauwels, V.R.N., Houser, P.R., Verhoest, N.E.C. (2006). Assimilation of point measurements to estimate spatial mean soil moisture. In preparation.

Chapter 2

Nature as a system, data as signals

2.1 Introduction

Measurements taken in nature can be seen as signals that are related to a natural system. The transformation of input signals to the resulting output signals happens within that system. Rainfall is an example of an input signal that causes soil moisture signals to change. The intrinsic properties of the land determine the characteristics of this change. In the next sections, some basics on signals and systems are reviewed, as a background for data analysis and modeling. Methods for system state estimation are discussed more in detail, using a notation that is based on the proposed unified notation of Ide *et al.* [1997]. Throughout this work, vectors are in lowercase bold and matrices in uppercase bold fonts.

2.2 Signals: data

Information in signals is captured in the pattern of changing values as function of a specific independent variable, e.g. time or space, in a discrete or a continuous sense. Deterministic signals are completely described by their mathematical functions. However, a lot of signals cannot be described by such functions and their description must be put in probabilistic terms: these signals are stochastic processes or random fields, i.e. a sequence (in time) or a field (in space) of random variables. Many signals are corrupted by noise, which is typically a random signal.

2.2.1 Description of random variables

Several authors [Papoulis, 1965; Jazwinski, 1970; Maybeck, 1979] gave a review of probability theory and its use in the characterization of signals. As opposed to discrete events (e.g. coin flipping), that are discussed in basic books on probability theory and described by discrete probability mass functions, the focus in this work is on so called continuous random vectors containing multiple variables, which are characterized by multi-dimensional probability distribution functions (=cumulative distribution function, cdf), or the derivatives, the frequency function (=probability density functions, pdf). The moments of these distributions completely describe the nature of the studied random variables.

2.2.2 Spatial and temporal characteristics

Pdfs and their moments describe the ensemble characteristics of signals, i.e. features of a random variable at a given time instant and a given point in space. For the analysis of observations, one is forced to study characteristics in space or time: ensemble information is not available. Analysis of observations used in this study will be performed by common methods used for time series analysis and by geostatistical tools in chapter 3. Besides the analysis in the time domain, which has been most common in hydrology, interesting information can also be obtained through an analysis in the frequency domain.

2.3 Deterministic systems

A system transforms input signals to output signals. The way this is done is a characteristic of the system. A model for a system can be simply defined as a representation of essential aspects of a physical system. To derive a mathematical model for a system, there exist mainly two methods:

- Theoretical analysis: dynamic properties are taken care of by the respective balance equations. After division of a process in subprocesses, laws of conservation of mass, energy and momentum are applied to the subsystems, resulting in a set of coupled linear or non-linear, ordinary or partial, differential or difference equations, which, together with the boundary conditions, form the mathematical model of the process.

- Experimental analysis (identification): a model type is proposed and parameters are estimated by surveying the behavior of the system in specific circumstances. If the structure of the model is known in advance, or at least can be assumed properly, parametric identification methods can be used. This will be illustrated in chapter 5, where a land surface model will be cali-

brated. If the model structure is not known, non-parametric procedures can be applied [De Keyser, 2001].

Models are classified in different ways: distributed parameter models versus lumped parameter models, discrete time (difference equations) versus continuous time models (differential equations), stochastic versus deterministic models, static versus dynamic models, parametric versus non-parametric models, linear versus non-linear models and input-output models versus state-space models. This will be further discussed for hydrological models in chapter 4. In this section, the basics of deterministic system models are reviewed as background and motivation for the discussion of stochastic models in section 2.4.

2.3.1 Deterministic discrete dynamical system

In practice, a model is based on several differential equations of different orders, which relate the input $\mathbf{u}$ (dimension p) to the output $\mathbf{y}$ (dimension m). These n-th order differential equations can be reduced to n first order equations. Together they compose a corresponding state-space representation of the system, being a first order vector differential equation. While in general, systems are introduced in continuous time, in this work only discrete models will be dealt with, because in most practical applications discrete-time models are used on digital computers and measurements are discrete. In the discrete case the state space representation is a first order difference equation.

Linear system

The deterministic discrete-time difference equation for a linear system model is given by:

$$\mathbf{x}_{i+1} = \mathbf{F}_{i+1,i}\mathbf{x}_i + \mathbf{B}_i\mathbf{u}_i \tag{2.1}$$

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x}_i \tag{2.2}$$

with $\mathbf{x}_i$ the n-dimensional state, $\mathbf{F}_{i+1,i}$ the $(n \times n)$-dimensional linear transition matrix, $\mathbf{u}_i$ the p-dimensional input to the model and $\mathbf{B}_i$ an $(n \times p)$-dimensional matrix relating the input to the state, in this case for a time-variant system. Expression (2.1) alone captures the system dynamics completely, but it is not a complete equivalent representation for general differential equations which relate the input to the output. The transformation of the state to the m-dimensional output or the measured signal $\mathbf{y}_i$ in Eq. (2.2) is performed by the matrix $\mathbf{H}_i$ of dimension $(m \times n)$. Eq. (2.1) is often referred to as the process model, while Eq. (2.2) is the measurement model.

The internal system variables composing the state of a system are of major importance for this study. In a state-space model, the inner variables are not eliminated,

as is usually done for input-output models. A state gives a means to characterize a system at a certain time with a minimum amount of variables. The influence of past time is completely included in the memory of the state. Together with the external influences or forcings on a system (e.g., precipitation and radiation for a land system), the states are sufficient to estimate the future behavior of the deterministic system. The order of a system is given by the amount of independent state variables in a state vector or the number of initial conditions. A certain set of state variables uniquely determines the system behavior, but there is an infinite number of such sets, e.g. physical, standard observable, standard controllable and canonical state space representations. Each matrix $\mathbf{F}$ can be transformed in an equivalent matrix by a similarity transformation, without changing the system dynamics. Some state space representations may be more suitable for estimation or control (section 2.3.2 and 2.4). This is determined by the observability and controllability [Kalman, 1960; Maybeck, 1979; Jazwinski, 1970] of a specific state space representation, rather than of the system itself. Controllability reflects on the effects of input on the states of the system, while observability is concerned with the effect of the model state on the outputs.

Non-linear system

For discrete non-linear systems the process and measurement equations are given by:

$$\mathbf{x}_{i+1} = \mathbf{f}_{i+1,i}(\mathbf{x}_i, \mathbf{u}_i) \tag{2.3}$$
$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i) \tag{2.4}$$

with $\mathbf{f}_{i+1,i}()$ the non-linear system function and $\mathbf{h}_i()$ the non-linear function relating the state to the output or measurements. These types of systems are often difficult to work with in an analytical way. Therefore, they are mostly linearized around a working point (for $\mathbf{x}_i = \mathbf{x}_i^*$ and $\mathbf{u}_i = \mathbf{u}_i^*$), leading to:

$$\mathbf{x}_{i+1} = \mathbf{f}_{i+1,i}(\mathbf{x}_i^*, \mathbf{u}_i^*) + \mathbf{F}_{i+1,i}[\mathbf{x}_i - \mathbf{x}_i^*] + \mathbf{B}_i[\mathbf{u}_i - \mathbf{u}_i^*] \tag{2.5}$$
$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i^*) + \mathbf{H}_i[\mathbf{x}_i - \mathbf{x}_i^*] \tag{2.6}$$

or the new linearized system process is now:

$$\mathbf{x}_{i+1} - \mathbf{x}_{i+1}^* = \mathbf{F}_{i+1,i}[\mathbf{x}_i - \mathbf{x}_i^*] + \mathbf{B}_i[\mathbf{u}_i - \mathbf{u}_i^*] \tag{2.7}$$

$\mathbf{F}_{i+1,i}$ and $\mathbf{B}_i$ are Jacobian matrices of dimensions $(n \times n)$ and $(n \times p)$ and are the derivatives of $\mathbf{f}_{i+1,i}$ to $\mathbf{x}$ and $\mathbf{u}$, evaluated in $\mathbf{x}_i^*$ and $\mathbf{u}_i^*$, respectively. $\mathbf{H}_i$ has

dimensions $(m \times n)$ and is the derivative of $\mathbf{h}_i$ to $\mathbf{x}$, evaluated for $\mathbf{x}_i^*$:

$$\mathbf{F}_{i+1,i} = \begin{bmatrix} \frac{\partial \mathrm{f}_1}{\partial x_1} & \cdots & \frac{\partial \mathrm{f}_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{f}_n}{\partial x_1} & \cdots & \frac{\partial \mathrm{f}_n}{\partial x_n} \end{bmatrix} |(\mathbf{x}_i^*, \mathbf{u}_i^*) \tag{2.8}$$

$$\mathbf{B}_i = \begin{bmatrix} \frac{\partial \mathrm{f}_1}{\partial u_1} & \cdots & \frac{\partial \mathrm{f}_1}{\partial u_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{f}_n}{\partial u_1} & \cdots & \frac{\partial \mathrm{f}_n}{\partial u_p} \end{bmatrix} |(\mathbf{x}_i^*, \mathbf{u}_i^*) \tag{2.9}$$

$$\mathbf{H}_i = \begin{bmatrix} \frac{\partial \mathrm{h}_1}{\partial x_1} & \cdots & \frac{\partial \mathrm{h}_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{h}_m}{\partial x_1} & \cdots & \frac{\partial \mathrm{h}_m}{\partial x_n} \end{bmatrix} |(\mathbf{x}_i^*) \tag{2.10}$$

where n, p and m are the dimensions of respectively the state, the input and output vector. The scalar element $x_{a,i+1} = \mathrm{f}_{a,i+1,i}(x_{1,i}, \ldots, x_{a,i}, x_{n,i}, u_{1,i}, \ldots, u_{b,i}, u_{p,i})$ and the scalar element $y_{c,i} = \mathrm{h}_{c,i}(x_{1,i}, \ldots, x_{a,i}, x_{n,i})$. $x_{a,i} \in \mathbf{x}_i$, with $a = 1, \ldots, n$, $u_{b,i} \in \mathbf{u}_i$ with $b = 1, \ldots, p$ and $y_{c,i} \in \mathbf{y}_i$ with $c = 1, \ldots, m$.

2.3.2 Deterministic system analysis and control

A feature of main importance for a deterministic system is its stability [Vansteenkiste and Van Welden, 1996; Peremans, 2000]. Control is often used as a means to remove the instability in systems or to push the system towards a desired behavior, without using too much energy. Control systems can be implemented in an open loop or a closed loop. In open loop the control is independent of the output, in contrast to closed loop systems with feedback. A lot of tools are developed to optimize the control of systems. For linear systems a widely used optimization criterium is a quadratic cost function [Willems, 2003] to penalize both the deviation of the state from a desired value and the used energy (forcings). This is known as the least squares (LQ) problem. A major optimization tool in modern control is H_2 optimization [Megretski, 2001], covering LQG (Least-Squares Gaussian) control and Kalman filtering.

Besides the wanted influence on the system by the control input $\mathbf{u}_i$, there is often a considerable impact of noise on the system. Further, measurements are taken to obtain knowledge about the system state. However, often the state is not measured directly and the measurements of the output are often contaminated by noise and/or only cover a part of the state. Finally, no mathematical model is perfect and the dynamical systems are not only driven by the input we know, but also by disturbances which we can neither control nor model deterministically. These factors highlight the need for stochastic control systems [Willems, 2003]. It can be shown that optimal control lays in a state estimation (or state reconstruction) followed by an optimal feedback of the estimated state (Separation Theorem). The

optimal (deterministic) control problem and optimal (stochastic) estimation are dual to each other: their solutions are obtained by solving analogous equations [Ghil and Malanotte-Rizzoli, 1991].

While for land systems, 'uncontrolled' meteorological forcings provide the input, the input to e.g. industrial processes is strongly controlled. In hydrological studies, the aim is to model processes as well as possible, only to obtain an as complete idea as possible of the system, without the intention to control nature directly on a short time basis. While in industry deviations from the desired behavior are simply not allowed by anticipating possible consequences on very short time scales through control, in hydrology mostly information from state estimation and predictions are used to deal with the consequences later, since it is impossible to control or regulate nature completely. The idea of regulation or control versus state estimation is illustrated in figure 2.1. The system enclosed by the upper solid box is the real system (i.e. how nature or an industrial process really works). In the dashed box a model is used for the state estimation procedure. In the state estimation procedure, we want to find the real state $\mathbf{x}$, which cannot be measured exactly nor modeled perfectly. The best estimate we can find is $\hat{\mathbf{x}}$. Once we know the state, this information can be used for monitoring and control. In control, the real physical system (e.g. a chemical process, a spacecraft) in the solid box will be influenced by feed-forward (FF) and feed-back (FB), when the behavior deviates from the desired one.

2.4 Stochastic systems and state estimation

2.4.1 Stochastic discrete dynamical systems: state propagation

In practice, systems are hardly completely deterministic and output (states) is typically a Markov process as result of a stochastic differential or difference equation

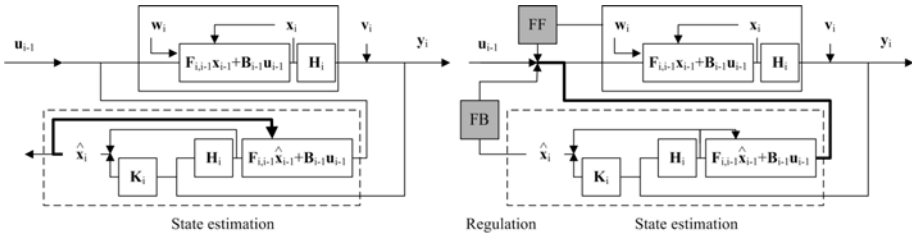


Figure 2.1: State estimation only versus control (including state estimation). The upper half parts represent the true linear system, while in the lower half parts the state estimation is shown (without and with feedback for control). The symbols are in the text discussing the Kalman filter.

(SDE or SΔE). A Markov process is a random process whose future probabilities are determined only by its most recent values. As stressed in many introductions in state estimation, e.g. by Maybeck [1979], deterministic system and control theory cannot help us sufficiently in a system analysis and controller design for real systems. The solution of stochastic discrete system equations is the main topic of this section. The focus will be on calculating the mean and (co)variances of stochastic processes, since in the normally distributed case, these quantities define the process completely. Higher moments are necessary when either the process is non-linear or the distribution of the state is non-Gaussian. The probability of a (static) random variable is constrained by a dynamical system, i.e. the dynamical system influences the evolution and shape of the state's pdf through time.

For measurements the deterministic approach is not very realistic either: in practice the relation between an observation $\mathbf{y}$ (m-dimensional) and a state $\mathbf{x}$ (n-dimensional) is not perfectly known and observations are always prone to some noise. In the following sections, with every system SΔE a corresponding measurement SΔE is given. However, it should be noted that system and measurement equations do not have to be of the same type: e.g., linear discrete measurements can be observations of a continuous non-linear system.

Discrete linear system

System description and solution
The discrete linear system, described by linear SΔE, is given by:

$$\mathbf{x}_{i+1} = \mathbf{F}_{i+1,i}\mathbf{x}_i + \mathbf{B}_i\mathbf{u}_i + \mathbf{G}_i\mathbf{w}_i \tag{2.11}$$

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x}_i + \mathbf{v}_i \tag{2.12}$$

with $\mathbf{w}_i$ a stochastic system or process noise, $\mathbf{G}_i$ an $(n \times n)$ matrix function which may color white noise $\mathbf{w}_i$, and $\mathbf{v}_i$ the observation noise. If the random input $\mathbf{w}_i$ would be absent, the expression for the system is an ordinary difference equation and $\mathbf{x}_i$ would simply be its solution. In the presence of random input $\mathbf{w}_i$, the pdf of $\mathbf{x}_i$, $f_{\mathbf{x}_i}(\boldsymbol{x})$ with $\boldsymbol{x}$ a realization of $\mathbf{x}$, is usually of interest, as it gives a complete description of the state $\mathbf{x}_i$.

Moments
As Gaussian density functions are preserved after linear transformation, only the first 2 moments should be calculated to describe the density function of $\mathbf{x}_i$ completely and they can be directly obtained from Eq. (2.11) with zero mean white Gaussian noise for $\mathbf{w}_i$, i.e. with $E[\mathbf{w}_i] = \mathbf{0}$ and $E[\mathbf{w}_i\mathbf{w}_i^T] = \mathbf{Q}_i$:

$$\hat{\mathbf{x}}_{i+1} = \mathbf{F}_{i+1,i}\hat{\mathbf{x}}_i + \mathbf{B}_i\mathbf{u}_i \tag{2.13}$$

$$\mathbf{P}_{i+1} = \mathbf{F}_{i+1,i}\mathbf{P}_i\mathbf{F}_{i+1,i}^T + \mathbf{G}_i\mathbf{Q}_i\mathbf{G}_i^T \tag{2.14}$$

with $\hat{\mathbf{x}}_i$ an estimate (e.g. the mean of the pdf) of the true $\mathbf{x}_i$ and $\mathbf{P}_i$ is the (co)variance matrix, reflecting the uncertainty of $\hat{\mathbf{x}}_i$.

Discrete non-linear system

System description and solution
Discrete non-linear systems are defined by:

$$\mathbf{x}_{i+1} = \mathbf{f}_{i+1,i}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{w}_i) \tag{2.15}$$
$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i, \mathbf{v}_i) \tag{2.16}$$

In the special case of additive Gaussian noise $\mathbf{w}_i$ and $\mathbf{v}_i$, these equations become:

$$\mathbf{x}_{i+1} = \mathbf{f}_{i+1,i}(\mathbf{x}_i, \mathbf{u}_i) + \mathbf{G}_i(\mathbf{x}_i)\mathbf{w}_i \tag{2.17}$$
$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i) + \mathbf{v}_i \tag{2.18}$$

Evolution of the pdf
The evolution of the probability density $f_{\mathbf{x}_{i+1}|\mathbf{x}_i}(\boldsymbol{x}_{i+1}|\boldsymbol{x}_i)$ of the Markov process generated by the non-linear discrete transition function can be found as described by Jazwinski [1970]. Most authors refer for simplicity to the evolution of the pdf by the Fokker-Planck or Kolmogorov's forward equation, which is valid for continuous non-linear systems. For non-linear processes the pdfs of the state cannot be described by Gaussian distributions and hence all moments should be calculated. Therefore, it is better advised to calculate the propagation of the pdf. Nevertheless, the first 2 moments give valuable information on the mean path and dispersion for linearized systems

It can be shown [Jazwinski, 1970] that the expression for the first moment for a non-linear model can be calculated as:

$$\hat{\mathbf{x}}_{i+1} = \mathbf{f}_{i+1,i}(\hat{\mathbf{x}}_i, \mathbf{u}_i) + \frac{1}{2}\boldsymbol{\mathcal{F}}''_{i+1,i}(\hat{\mathbf{x}}_i, \mathbf{u}_i)vec[\mathbf{P}_i] \tag{2.19}$$

with $\boldsymbol{\mathcal{F}}''_{i+1,i}(\hat{\mathbf{x}}_i, \mathbf{u}_i)$ the Hessian matrix of the system function $\mathbf{f}_{i+1,i}()$ and $vec[\cdot]$ standing for a vector constructed by concatenation of the different columns in a matrix in one vector. It is clear from this last expression that the mean is dependent of the variance, which is one of the many unpleasant properties of non-linear systems. The expression for $\mathbf{P}_{i+1}$ (Eq. (2.14)) can be used as an approximation for the discrete non-linear systems after linearization, by replacing $\mathbf{F}_{i+1,i}$ by the Jacobian matrix of the non-linear function. Note again that the mean and covariance do not describe the pdf completely here.

2.4.2 Corrected estimation using observations: state update

In the absence of observations, the propagation of the first 2 moments was discussed in subsection 2.4.1. Estimating the state of a stochastic dynamical system when noisy observations are available is a problem that dates back to the work of Gauss in 1795 [Sorenson, 1970]. He developed the technique of least squares

for orbit determination. The problem was looked at as a deterministic problem of minimizing errors.

The optimal estimate is sought as a combination of measurement data, information from a system model and statistical description of uncertainties. The estimation problem can be subdivided in filtering, smoothing and prediction problems. Given a number of observations in time $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{i+k}\}$, the estimation problem consists of computing the estimate of $\mathbf{x}_i$. If $k > 0$, then we are dealing with a smoothing problem, if $k < 0$, it is a prediction problem and in case $k = 0$, it is a filtering problem. In this work, the focus will be on estimation of the complete state. However, sometimes it is convenient not to solve the problem for all components in the model state, but for a reduced control vector.

In the next sections, we will focus on statistical methods for state estimation through filtering, without a full probabilistic (or Bayesian) interpretation [Jazwinski, 1970; Maybeck, 1979; Lorenc, 1995; van Leeuwen and Evensen, 1996]. By the latter approach, the conditional probability density function of the state, given the measurements, is derived based on Bayes' rule and the optimal (minimum variance) estimate $\hat{\mathbf{x}}_{k|l}$ is given by the conditional mean (which depends on higher-order moments in non-linear problems). The statistical approach does not involve stochastic differential/difference equations, which makes the results conceptually and theoretically simpler. Statistical methods (e.g. least squares, maximum likelihood,...) are typically developed for linear estimation problems [Cohn, 1997] and formally applied to the non-linear estimation problem, e.g. after linearization.

Observability, controllability
As mentioned earlier, the choice of the state space representation determines the controllability and observability of the system. The number of observations required for the determination of unknown quantities is function of the observability of the system and of major importance for state estimation [Cohn and Dee, 1988].

Conditional mean/mode and (non-)linearity
The essence of state estimation is the description of the state by a statistic that represents the state's conditional density function. The conditional mean estimation has mostly been preferred, because it is a minimum variance estimate, independent of the distribution of the stochastic forcings $\mathbf{w}_i$ and $\mathbf{v}_i$. However, one could also estimate the conditional mode, which is a maximum a posteriori probability estimation. These are equal for Gaussian statistics, and differences are mainly due to non-linearity. A least-squares analysis can be applied to any problem, but it will not yield optimal results if the underlying distributions are not Gaussian. For non-linear dynamics, there is no way to assume that an initially Gaussian state error distribution would remain Gaussian after propagation. However, as will be illustrated, several techniques designed for estimation under linear dynamics, have been applied for non-linear problems, after linearization of the model dynamics, often yielding unsatisfying results in highly non-linear problems [Miller *et al.*, 1994]. In the remainder of this work, the focus will be on linear discrete filters, but continuous extensions exist [Jazwinski, 1970]. Miller *et*

al. [1999] suggested that better results can be obtained by using non-linear filters, which are promising, but computationally intensive. Statistics directly calculated from the estimated probability density function of the system's state, conditioned on the observations, then yield the assimilation products.

2.5 Sequential assimilation

Through sequential methods for assimilation, a correction is added to an a priori (forecasted) state estimate $\hat{\mathbf{x}}_i^-$ to obtain an a posteriori (analysis) estimate $\hat{\mathbf{x}}_i$ of the true state $\mathbf{x}_i$. These methods are sometimes referred to as direct observer assimilation methods and are mostly intermittent. The a priori state estimate $\hat{\mathbf{x}}_i^-$, also referred to as background, can be a climatology or a trivial state, a prediction from a forecast model or the output from a previous analysis. The linear combination is generally given by:

$$\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \mathbf{K}_i[\mathbf{y}_i - \mathbf{h}_i(\hat{\mathbf{x}}_i^-)] \tag{2.20}$$

with $\mathbf{K}_i$ a weighting factor. For earth systems, the state vector generally consists of a number of variables at different locations in space. A logical feature of data assimilation is that in data rich areas the analysis is dominated by the observations, while in data poor regions the a priori estimates dominate, but they are improved by the information from surrounding areas. For distributed models, forecasts are able to propagate information from data rich to data poor areas (in case these areas are linked somehow through the model). In this chapter, mainly the technical details of different assimilation schemes are reviewed. For the discussion of practical applications, the reader is referred to chapter 7.

2.5.1 Direct insertion/replacement

As is common in practice, observations are assumed to provide the best estimates of a state and a priori state estimates by the model are simply replaced by the corresponding observations. It can be seen as a linear combination (Eq. (2.20)) of the a priori estimate and the a priori residual, weighted by $\mathbf{K}_i = \mathbf{I}$, the unity matrix. The risk of this approach is that unbalanced state estimates may result, which causes model shocks: the model will attempt to restore the dynamic balance that would have existed without insertion. The model needs to propagate information towards unobserved parts of the system. Some applications of this method were described e.g. by Gauntlett and Seaman [1974], Li and Islam [1999] and Heathman *et al.* [2003].

2.5.2 Statistical correction

Closely related to the method of direct insertion is assimilation by statistical correction [Houser *et al.*, 1998]. This method adjusts the mean and variance of the model states to match the statistics describing the observations. The model needs to propagate information towards unobserved parts of the system. Some practical applications are described e.g. by Pauwels *et al.* [2001] and Pauwels *et al.* [2002].

2.5.3 Successive correction

The successive corrections method (SCM) was developed by Bergthorsson and Doos [1955] and Cressman [1959], and is also known as observation nudging. The scheme begins with an a priori state estimate (background field) for an individual (scalar) variable $\hat{x}_i \in \hat{\mathbf{x}}_i$, which is successively adjusted by nearby observations in a series of scans (iterations, k) through the data. The analysis at time step i is found by passing through following sequence of updates:

$$\begin{cases} \hat{x}_i^0 & = & \hat{x}_i^- \\ \hat{x}_i^1 & = & \hat{x}_i^0 + \mathbf{k}_i^{1^T}[\mathbf{y}_i - \mathbf{h}_i(\hat{\mathbf{x}}_i^0)] \\ & \vdots & \\ \hat{x}_i^{k+1} & = & \hat{x}_i^k + \mathbf{k}_i^{k^T}[\mathbf{y}_i - \mathbf{h}_i(\hat{\mathbf{x}}_i^k)] \end{cases} \tag{2.21}$$

with $\mathbf{h}_i(\hat{\mathbf{x}}_i^k)$ the value of the state estimate at the kth iteration, evaluated at the m observation points ($\mathbf{h}_i$ is the interpolation operator), $\mathbf{y}_i$ the vector of observations within a predefined influence radius R_i^k and $\mathbf{k}_i^k$ is a vector of weights for all observations within the predefined radius of influence. The m elements of $\mathbf{k}_i^k$ are given by:

$$k_{j,i}^k = \frac{c_{j,i}^k}{q^2 + \sum_{j=1}^m c_{j,i}^k} \tag{2.22}$$

with q an estimate of the ratio of the observation error to the background error covariance, $c_{j,i}^k$ any sort of weights and $j = 1, \ldots, m$. For the Cressman scheme, the observations are assumed to be perfect ($q^2 = 0$) and the weights are given by:

$$c_{j,i}^k = \begin{cases} \frac{{R_i^k}^2 - d_j}{{R_i^k}^2 + d_j^2} & \text{for} \quad d_j \leq R_i^k \\ 0 & \text{for} \quad d_j > R_i^k \end{cases} \tag{2.23}$$

with R_i^k the radius of influence, which is mostly shrinking for successive iterations k, so that the field is corrected to larger scale features during the first iterations, and conforms to smaller scale features during later iterations and d_j the distance between the jth observation point and the grid point for the analysis. Different weighting functions could be proposed, such as a Gaussian function [Barnes, 1964], where $c_{j,i}^k = \exp[-d_j^2/(2{R_i^k}^2)]$.

If the estimation of the complete state vector $\hat{\mathbf{x}}_i$ would be considered, for each iteration k, an equation $\hat{\mathbf{x}}_i^{k+1} = \hat{\mathbf{x}}_i^k + \mathbf{K}_i^k[\mathbf{y}_i - \mathbf{h}_i(\hat{\mathbf{x}}_i^k)]$ would be obtained with $\mathbf{K}_i$ a matrix containing an empirically derived weighting, that takes into account the spatial distribution of observations.

The advantage of this method lies in its simplicity and mostly satisfying results. However, in case of observational error or different sources (and accuracies) of observations, this scheme is not a good option for assimilation, since information on the observation accuracy is not accounted for. Further, the radii of influence are user-defined and should be determined by trial and error or more sophisticated methods that reduce the advantage of its simplicity. The weighting functions are empirically chosen and are not derived based on physical or statistical properties. Obviously, this method is not effective in data sparse regions. Some practical examples are discussed by Bratseth [1986] and Daley [1991].

2.5.4 Analysis correction

The analysis correction method is a modification to the successive correction approach, in which the observation vector is also successively updated [Lorenc *et al.*, 1991]. In practice, assumptions to simplify this method result in an update equation which is equivalent to the one for optimal interpolation (see section 2.5.6).

2.5.5 Nudging

Nudging [Hoke and Anthes, 1976] or Newtonian relaxation consists of adding a term to the prognostic model equations that causes the solution to be gradually relaxed towards the observations. Nudging is very similar to the successive corrections technique and only differs in the fact that through the numerical model the time dimension is included [Stauffer and Seaman, 1990]. Again, to allow some link with most literature on this method, first the equations are given for the update of a scalar variable x (which belongs to a vector $\mathbf{x}$) for a continuous assumed system:

$$\frac{\partial x}{\partial t} = f(x(\mathrm{x},\mathrm{y},\mathrm{z},t),\mathrm{x},\mathrm{y},\mathrm{z},t) + \mathbf{k}(\mathrm{x},\mathrm{y},\mathrm{z},t)^T[\mathbf{y}_i - \mathbf{h}_i(\hat{\mathbf{x}}^-(\mathrm{x},\mathrm{y},\mathrm{z},t))] \qquad (2.24)$$

The vector $\mathbf{k}(\mathrm{x},\mathrm{y},\mathrm{z},t)$ contains elements which all consist of (1) a nudging coefficient, which determines the magnitude of the correction term relative to all models processes and is equal for all elements within the vector and only dependent on the variable x to be analyzed, (2) a weighting function, which determines the spatial and temporal variation of the nudging coefficient, depends on the variable x and is function of space $\mathrm{x},\mathrm{y},\mathrm{z}$ and time t and (3) an observational quality factor varying between 0 and 1, based on the quality of the data which contributed to the analysis. The operator $\mathbf{h}_i$ is introduced to select those state variables from the state vector that correspond directly to the observations (no interpolation) or to interpolate the state variables to the locations of the observations.

Two distinct approaches have been developed [Stauffer and Seaman, 1990; Stauffer *et al.*, 1991]. In the first method (*analysis nudging*), the nudging term for a given variable is proportional to the difference between the model simulation at a given grid point and an 'analysis' of observations (i.e. processed observations) calculated at the corresponding grid point. In this case $\mathbf{k}(\mathrm{x},\mathrm{y},\mathrm{z},t)$, $\mathbf{y}_i$ and $\mathbf{h}_i(\hat{\mathbf{x}}^-(\mathrm{x},\mathrm{y},\mathrm{z},t))$ reduce to a scalars in Eq. (2.24). The scalar $k(\mathrm{x},\mathrm{y},\mathrm{z},t)$ can then be written as:

$$k(\mathrm{x},\mathrm{y},\mathrm{z},t) = g_x c_{\mathrm{x,y,z},t} q_{\mathrm{x,y,z}} \tag{2.25}$$

with g_x the nudging coefficient, $c_{\mathrm{x,y,z},t}$ the four-dimensional weighting function and $q_{\mathrm{x,y,z}}$ a quality factor. The scalar y_i is an interpolated observational value and the operator $\mathbf{h}_i$ selects those state variables from the state vector that correspond directly to the observations (no interpolation).

In the second method (*observation nudging*), the difference between the model simulation and the observed state is calculated at the observation locations. To update a single variable, m observations within a preset radius are included, i.e. all vectors in the nudging term in Eq. (2.24) have dimension m. The vector $\mathbf{k}(\mathrm{x},\mathrm{y},\mathrm{z},t)$ now contains elements $k_j(\mathrm{x},\mathrm{y},\mathrm{z},t)$ given by:

$$k_j(\mathrm{x},\mathrm{y},\mathrm{z},t) = g_x \frac{c^2_{j,\mathrm{x,y,z},t} q_j}{\sum_{j=1}^{m} c_{j,\mathrm{x,y,z},t}} \tag{2.26}$$

with j referring to the jth observation, g_x the nudging coefficient, $c_{j,\mathrm{x,y,z},t}$ the four-dimensional weighting function and q_j a quality factor. The vector $\mathbf{y}_i$ consists of an array of direct observations and $\mathbf{h}_i$ is now an interpolation operator.

For analysis nudging, $c_{\mathrm{x,y,z},t}$ is generally taken to be 1. For observation nudging, the weighting function $c_{\mathrm{x,y,z},t}$ for each observation is a combination of a horizontal, vertical and a temporal weighting function, given by c_{xy}, c_{z} and c_t, respectively. Suggestions for these weighting functions have been reviewed by Houser *et al.* [1998]. For example, for c_{xy} a Cressman type horizontal weighting could be used.

Eq. (2.24) can be written in discrete form for a vector state as follows:

$$\hat{\mathbf{x}}_i = \mathbf{f}_{i,i-1}(\hat{\mathbf{x}}_{i-1}) + \sum_{k=0}^{f(c_t)} \mathbf{k}^T_{i,i-k}[\mathbf{y}_{i-k} - \mathbf{h}_{i-k}(\hat{\mathbf{x}}^-_{i-k})] \tag{2.27}$$

with $\mathbf{k}^T_{i,i-k}$ an expression for $\mathbf{k}(\mathrm{x},\mathrm{y},\mathrm{z},t)^T\Delta t$ and Δt the discrete time step. This method basically uses observations at any time step $i-k$ ($\forall 0 \leq k \leq f(c_t)$, with $f(c_t)$ a limit imposed by the temporal weighting function c_t) to cause an update at next time steps i, which differs from the most current filtering methods in earth sciences, but which is closer to predictive filtering (see below).

In case Eq. (2.24) is extended to estimate complete state vectors, instead of scalars, the vector $\mathbf{k}(t)$ should be replaced by a gain matrix $\mathbf{K}(t)$. This gain matrix is then approximated by an empirical function containing a nudging factor, an observational quality factor and a temporal and spatial weighting function. In case

of analysis nudging, $\mathbf{K}(t)$ is a diagonal matrix. Some examples of applications of this method are described by Stauffer *et al.* [1991], Stauffer and Seaman [1994], Seaman *et al.* [1995], Houser *et al.* [1998] and Paniconi *et al.* [2003].

2.5.6 Optimal interpolation (OI)

The optimal interpolation or statistical interpolation method [Daley, 1991] is an approximation of the Kalman filter (see next section). It is also a least squares method. This method differs from successive correction methods and nudging methods in the fact that the observations are weighted according to some known or estimated error statistics, rather than by empirical values. This technique is closely related to kriging. For this method, as for all methods discussed above, observations are mostly selected only in a limited geometrical domain around the model variable to be estimated, often by imposing small a priori error covariances for large separation. When the a priori error covariance matrix is assumed to be static, it is most often named the background error covariance.

The structure of $\mathbf{K}_i$ is built in the same way as for the optimal Kalman filter, but the a priori state error covariance matrix $\mathbf{P}_i^-$ (cfr. infra) is approximated by $\mathbf{P_S}_i\mathbf{P_C}\mathbf{P_S}_i$, with $\mathbf{P_S}_i$ a diagonal matrix of forecast error standard deviations and $\mathbf{P_C}$ the time-invariant correlation matrix, which is often defined dependent on the distance between points. The determination of $\mathbf{P_C}$ is a major issue for this technique and will be revisited in chapter 9. While for the above methods, no propagation of the uncertainty was included, the a posteriori error variances $\mathbf{P_S}_i$ can be propagated in a simple way, while the correlation matrix $\mathbf{P_C}$ is kept static. Some examples of this method are discussed by Houser *et al.* [1998] and Lorenc [1981].

2.5.7 Linear discrete Kalman filter (KF)

The Kalman filter yields a Best Linear Unbiased Estimator (BLUE), with a value for $\mathbf{K}$ that yields an optimal estimate [Jazwinski, 1970; Maybeck, 1979; Welch and Bishop, 2003; Brown and Hwang, 1992; Willems, 2003]. Because the KF will be used later in this work, a simple statistically based but founded derivation of its equations is provided, which will highlight the essential underlying assumptions. Compared to the above methods, the KF can be defined as an adaptive filter, for which the different factors alter with time or with the evolving state (see also chapter 9). The naming 'filter' refers to the fact that the KF is a low pass filter which removes noise from the observations.

Consider a discrete system, described by the linear stochastic difference equations given by Eq. (2.11) and (2.12).

$$\mathbf{x}_i = \mathbf{F}_{i,i-1}\mathbf{x}_{i-1} + \mathbf{B}_{i-1}\mathbf{u}_{i-1} + \mathbf{w}_i \tag{2.28}$$

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x}_i + \mathbf{v}_i \tag{2.29}$$

with vector $\mathbf{x}_i$, $\mathbf{y}_i$ and $\mathbf{u}_i$ defined earlier and with dimension n, m and p, respectively. The random vector $\mathbf{w}_i$ of dimension n represents process noise and the vector $\mathbf{v}_i$ of dimension m is measurement noise. The process noise can include input errors (e.g. error on the forcing measurements) as well as system errors because of the incomplete simplified representation of the system. These random noise vectors are assumed to be white and zero mean Gaussian and independent $\forall i$ with $\forall k \neq 0$:

$$E[\mathbf{w}_i] = \mathbf{0} \quad , \quad E[\mathbf{v}_i] = \mathbf{0}, \tag{2.30}$$

$$E[\mathbf{w}_i \mathbf{w}_{i+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{w}_i \mathbf{w}_i^T] = \mathbf{Q}_i, \tag{2.31}$$

$$E[\mathbf{v}_i \mathbf{v}_{i+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{v}_i \mathbf{v}_i^T] = \mathbf{R}_i, \tag{2.32}$$

$$E[\mathbf{w}_i \mathbf{v}_{i+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{w}_i \mathbf{v}_i^T] = \mathbf{0} \tag{2.33}$$

$\mathbf{Q}_i$ and $\mathbf{R}_i$ are positive definite covariance matrices describing the process and measurement noise. For clarity all variables involved in the filter to be developed, are explained in table 2.1. It is assumed that the system is completely observable and controllable, i.e.:

$$rank([\mathbf{H}^T \ \ (\mathbf{HF})^T \ \cdots \ (\mathbf{HF}^{n-1})^T]^T) \ = \ n \tag{2.34}$$

$$rank([\mathbf{B} \ \ \mathbf{FB} \ \cdots \ \mathbf{F}^{n-1}\mathbf{B}]) \ = \ n \tag{2.35}$$

We want to find a best estimate for the state vector $\mathbf{x}_i$. There are several ways to quantify the quality of the estimation of $\mathbf{x}_i$, written as $\hat{\mathbf{x}}_i$. A typical criterium is the least squares approximation, stating that the best estimate $\hat{\mathbf{x}}_i$ is found when the conditional expectation $E[(\mathbf{x}_i - \hat{\mathbf{x}}_i)^T \mathbf{M}_i (\mathbf{x}_i - \hat{\mathbf{x}}_i) | \text{available information}]$ is minimal for any $\mathbf{M}_i \geq 0$. All available output measurements and a priori statistics of $\mathbf{x}_i$ are considered as available information.

Estimation by instantaneous filtering

If an optimal estimate $\hat{\mathbf{x}}_i$ is calculated using all available observations at time steps $i - k \leq i$ ($\forall k \geq 0$), then the estimation procedure is filtering sensu stricto. The state is estimated (a posteriori estimate $\hat{\mathbf{x}}_i$) as an optimal combination of the a priori estimated state $\hat{\mathbf{x}}_i^-$ and the observation $\mathbf{y}_i$ measured at the same time (the procedure is waiting for the measurement), taking into account their relative accuracy, which is included in the error covariance matrices and end up in the weighting factor $\mathbf{K}_i$:

$$\boxed{\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \mathbf{K}_i[\mathbf{y}_i - \mathbf{H}_i \hat{\mathbf{x}}_i^-]} \tag{2.36}$$

The factor $\mathbf{K}_i[\mathbf{y}_i - \mathbf{H}_i \hat{\mathbf{x}}_i^-]$ is called the analysis increment. The forecasted (model) a priori estimate $\hat{\mathbf{x}}_i^-$ is found by propagating $\hat{\mathbf{x}}_{i-1}$ to $\hat{\mathbf{x}}_i^-$ by the system difference equation $\hat{\mathbf{x}}_i^- = \mathbf{F}_{i,i-1} \hat{\mathbf{x}}_{i-1} + \mathbf{B}_{i-1} \mathbf{u}_{i-1}$. The $\mathbf{K}_i$-matrix is found by minimizing the a posteriori estimate error covariance matrix. The a posteriori estimate error $\mathbf{e}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$ should be minimal, with $\mathbf{e}_i$ given by:

$$\mathbf{e}_i = \mathbf{e}_i^- - \mathbf{K}_i \mathbf{H}_i \mathbf{e}_i^- - \mathbf{K}_i \mathbf{v}_i \tag{2.37}$$

Table 2.1: Description of all variables involved in the description of the Kalman filter.

$\mathbf{x}_i$	True state vector	(n)
$\hat{\mathbf{x}}_i$	A posteriori estimated state vector	(n)
$\hat{\mathbf{x}}_i^-$	A priori estimated state vector	(n)
$\mathbf{y}_i$	Observation vector	(m)
$\mathbf{u}_i$	Control input vector	(p)
$\mathbf{w}_i$	Process noise vector	(n)
$\mathbf{v}_i$	Observation noise vector	(m)
$\mathbf{F}_{i+1,i}$	State transition matrix relating state $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$	$(n \times n)$
$\mathbf{f}_{i+1,i}$	Non-linear system dynamics relating state $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$	(n)
$\mathbf{B}_i$	Matrix relating control input $\mathbf{u}_i$ to state $\mathbf{x}_{i+1}$	$(n \times p)$
$\mathbf{H}_i$	Matrix relating state $\mathbf{x}_i$ to observation $\mathbf{y}_i$	$(m \times n)$
$\mathbf{K}_i$	Kalman gain = matrix relating residual $[\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_i]$ to state $\mathbf{x}_{i+1}$ or $\mathbf{x}_i$	$(n \times m)$
$\mathbf{e}_i^-$	A priori estimation error = $\mathbf{x}_i - \hat{\mathbf{x}}_i^-$	(n)
$\mathbf{e}_i$	A posteriori estimation error = $\mathbf{x}_i - \hat{\mathbf{x}}_i$	(n)
$\mathbf{P}_i^-$	A priori estimation error covariance matrix	$(n \times n)$
$\mathbf{P}_i$	A posteriori estimation error covariance matrix	$(n \times n)$
$\mathbf{R}_i$	Measurement error covariance matrix	$(m \times m)$
$\mathbf{Q}_i$	System error covariance matrix	$(n \times n)$
$\mathbf{r}_i^-$	Innovation or a priori residual = $\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_i^-$	(m)
$\mathbf{r}_i$	Residual = $\mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_i$	(m)

using Eq. (2.36) and (2.29), and with $\mathbf{e}_i^- = \mathbf{x}_i - \hat{\mathbf{x}}_i^-$ the a priori estimation error. Consequently, matrix $\mathbf{P}_i = E[\mathbf{e}_i\mathbf{e}_i^T]$ can be expressed as:

$$\mathbf{P}_i = [\mathbf{I} - \mathbf{K}_i\mathbf{H}_i]\mathbf{P}_i^-[\mathbf{I} - \mathbf{K}_i\mathbf{H}_i]^T + \mathbf{K}_i\mathbf{R}_i\mathbf{K}_i^T \tag{2.38}$$

The a priori estimate error covariance matrix (assuming zero mean of the error) is given by:

$$\mathbf{P}_i^- = E[(\mathbf{F}_{i,i-1}\mathbf{e}_{i-1} + \mathbf{w}_i)(\mathbf{F}_{i,i-1}\mathbf{e}_{i-1} + \mathbf{w}_i)^T] \tag{2.39}$$

Thus $\mathbf{P}_i^-$ can be found as:

$$\mathbf{P}_i^- = \mathbf{F}_{i,i-1}\mathbf{P}_{i-1}\mathbf{F}_{i,i-1}^T + \mathbf{Q}_i \tag{2.40}$$

which is exactly the expression (2.14) for the propagation of the covariance matrix in case $\mathbf{G} = \mathbf{I}$, with $\mathbf{I}$ a square matrix with 1 on the diagonal and 0 off-diagonal and when the estimate corresponds to the (conditional) mean of the true state. Reorganization and grouping of all terms (linear and quadratic) in $\mathbf{K}_i$ for Eq. (2.38) yields:

$$\begin{aligned}\mathbf{P}_i = (\mathbf{K}_i - \mathbf{P}_i^-\mathbf{H}_i^T[\mathbf{R}_i + \mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T]^{-1})\left[\mathbf{R}_i + \mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T\right] \\ (\mathbf{K}_i - \mathbf{P}_i^-\mathbf{H}_i^T[\mathbf{R}_i + \mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T]^{-1})^T \\ + \mathbf{P}_i^- - \mathbf{P}_i^-\mathbf{H}_i^T[\mathbf{R}_i + \mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T]^{-1}\mathbf{H}_i\mathbf{P}_i^-\end{aligned} \tag{2.41}$$

Consequently, the a posteriori estimation error $E[\mathbf{e}_i^T\mathbf{M}_i\mathbf{e}_i] = tr(\mathbf{M}_i\mathbf{P}_i)$ ($tr()$ is the sum of the elements on the diagonal, i.e. the trace) is minimized for the optimal Kalman gain $\mathbf{K}_i$:

$$\boxed{\mathbf{K}_i = \mathbf{P}_i^-\mathbf{H}_i^T[\mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T + \mathbf{R}_i]^{-1}} \tag{2.42}$$

and $\mathbf{P}_i$ may now be computed with this optimal gain $\mathbf{K}_i$ and Eq. (2.40) as:

$$\mathbf{P}_i = \mathbf{P}_i^- - \mathbf{P}_i^- \mathbf{H}_i^T [\mathbf{H}_i \mathbf{P}_i^- \mathbf{H}_i^T + \mathbf{R}_i]^{-1} \mathbf{H}_i \mathbf{P}_i^- \quad (2.43)$$

$$\mathbf{P}_i = \mathbf{P}_i^- - \mathbf{K}_i \mathbf{H}_i \mathbf{P}_i^- \quad (2.44)$$

with initial condition $\mathbf{P}_0^- = E[\mathbf{x}_0^- \mathbf{x}_0^{-T}] - E[\mathbf{x}_0^-]E[\mathbf{x}_0^-]$, the zero state error covariance matrix, covering all knowledge about the process prior to $i = 0$.

Expression (2.38) is valid for any gain $\mathbf{K}_i$. Eqs. (2.43) and (2.44) are valid only for the optimal Kalman [Kalman, 1960] gain $\mathbf{K}_i$, as terms going to zero under optimal estimation are eliminated in Eq. (2.41) in that case and only terms that cannot be eliminated by choosing an appropriate $\mathbf{K}_i$ make up the resulting expression.

The above derivations are valid for $\mathbf{w}_i$ and $\mathbf{v}_i$ white noise senso strictu (stochastic independence) as well as white noise in broad sense (no correlation). Statistical independence of $\mathbf{w}_i$ and $\mathbf{v}_{i+k}$ is essential for the validity of this analysis. However, statistical independence of $\mathbf{w}_i$ and $\mathbf{v}_i$ is only needed for simplicity of the mathematical analysis.

There are numerous other ways to find the expression for the optimal Kalman gain, as illustrated by Jazwinski [1970].

Filtering versus real-time predictive filtering

For real-time applications, the above derivation of the KF is not well suited, since it requires that observations are taken at the time of estimation and predictions cannot be performed until the observations reached the filtering algorithm. Therefore, in real-time applications a time delay between the measurement of observations and the use for state estimation is built into the KF procedure. Most applications in engineering use this approach [Willems, 2003].

Remarks

- The KF is an optimal recursive filter, that does not need all previous data in storage to be reprocessed every time a new measurement is available. Per update, all useful information is extracted from the observations. This makes the algorithms far more appealing than the equivalent Wiener filter [Willems, 2003; Maybeck, 1979].
- The Kalman-Bucy filter gives the equation for the Best Linear Unbiased Estimator $\hat{\mathbf{x}}_i$ (BLUE), described by the conditional mean and the conditional covariance matrix $\mathbf{P}_{i|i-k}$ of the Gaussian conditional density function. In a review on estimation theory, Cohn [1997] repeated that the conditional mean is always unique and the minimum (error) variance estimate, independently of the nature of pdfs of the errors in the various information sources. It can be shown that for linear problems this $\mathbf{P}_{i|i-k}$ is independent of the observations

and therefore also the unconditional covariance matrix $\mathbf{P}_i$, which may be precomputed in advance.

- *Kalman gain:* $\mathbf{K}_i$
  From the equation for the Kalman gain, it is clear that the residual weighs more heavily if the uncertainty on the measurements is low. At the other hand $\mathbf{K}_i$ will be very small for small $\mathbf{P}_i^-$, meaning that the model forecasts are quite certain and do not need correction.

- *Error covariance matrices:* $\mathbf{Q}_i$ *and* $\mathbf{R}_i$
  Under the conditions where $\mathbf{Q}_i$ and $\mathbf{R}_i$ are in fact constant, the estimation error covariance $\mathbf{P}_i$ and the Kalman gain $\mathbf{K}_i$ will stabilize quickly and remain constant and could be precomputed. Tuning the filters parameters $\mathbf{Q}_i$ and $\mathbf{R}_i$ may result in a superior filter (see adaptive filtering in chapter 9).

- *State error covariance matrix:* $\mathbf{P}_i$
  The a priori estimation error covariance $\mathbf{P}_{i+1}^-$ can be thought of as the covariance matrix, conditioned on the realizations of the observations $\mathbf{y}_i$, while the a posteriori estimation error covariance $\mathbf{P}_{i+1}$ is conditioned on the observations $\mathbf{y}_{i+1}$ in the filtering (sensu stricto) case. The expressions for $\mathbf{P}_i$ can be put in many different forms as given by e.g. Maybeck [1979].

  The elements of $\mathbf{P}_i$ describe the spatial structure of the error, while its norm, given by $tr(\mathbf{P}_i)$, provides a quantification of this error. $\sqrt{tr(\mathbf{P}_i)}$ gives the norm of the estimation error vector, which was to be minimized.

- *Innovations*
  Under the assumption of zero mean error, the innovations sequence $\mathbf{r}_i^- = \mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_i^-$ is zero-mean white Gaussian for linear models. $\mathbf{r}_i^-$ is independent of previous $\mathbf{r}_{i-k}^-$ $(\forall k > 0)$ with

$$E[\mathbf{r}_i^-] = \mathbf{0} \tag{2.45}$$

$$E[\mathbf{r}_i^- {\mathbf{r}_i^-}^T] = \mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T + \mathbf{R}_i \tag{2.46}$$

  Analysis of the a priori residuals during filtering can reveal filtering failures, e.g. for consistent high innovations [Maybeck, 1979]. Note that sometimes the innovations are called residuals. In this work, the term residuals is reserved for analysis residuals $\mathbf{r}_i = \mathbf{y}_i - \mathbf{H}_i\hat{\mathbf{x}}_i$, while a priori residuals is a synonym for innovations.

- *Gaussian error assumption and optimality*
  Cohn [1997] indicated clearly that the KF relies on Gaussian distributed errors ($\mathbf{w}_i$ and $\mathbf{v}_i$) with zero mean for the analysis step. No Gaussian assumptions are needed for propagation of the state and its uncertainty. With non-Gaussian error distributions, the linear KF does not yield a conditional mean as a minimum variance analysis, but it is still optimal in the class of linear filters, i.e. the estimation error is not correlated to the measurements. For Gaussian errors, the KF is optimal in the class of both linear and non-linear filters, i.e. the estimation error is independent of the observations.

- *Smoothing or filtering with retardation*
  For some applications, the state $\hat{\mathbf{x}}_i$ is to be estimated, given the observations $\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{i+k}$ for $k \geq 0$. In case $k = 0$, the discussion on filtering (cfr. supra) solves the problem. For $k > 0$, the problem is more complicated, as the extra available information includes a stochastic component that is depending on earlier information. Smoothing is classified in three categories: fixed-interval smoothing, fixed-point smoothing and fixed-lag smoothing [Brown and Hwang, 1992].

2.5.8 Extended Kalman filter (EKF)

Consider a discrete system, described by the non-linear stochastic difference equations for the system dynamics and measurements:

$$\mathbf{x}_i = \mathbf{f}_{i,i-1}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1}, \mathbf{w}_{i-1}) \tag{2.47}$$
$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i, \mathbf{v}_i) \tag{2.48}$$

The system and observation error $\mathbf{w}_i$ and $\mathbf{v}_i$ are now not necessary of dimension n and m, as defined for the KF for linear systems with additional noise, but of any dimension r and s, respectively. There are two possibilities to linearize the system equation: (i) a linearization about some nominal trajectory in state-space that does not depend on the measurement data (=*linearized Kalman filter*) or (ii) linearization about the trajectory that is continually updated with the state estimates, calculated using measurements (=*extended Kalman filter*). The main difference is the way the Jacobian matrices of the partial derivatives are determined. Once these are known, the procedure is completely analogous to the basic KF. Based upon the Taylor hypothesis, the state $\mathbf{x}_i = \hat{\mathbf{x}}_i^- + \Delta\mathbf{x}_i$, with $\hat{\mathbf{x}}_i^- = \mathbf{f}_{i,i-1}(\hat{\mathbf{x}}_{i-1}, \mathbf{u}_{i-1}, \mathbf{0})$ and $\Delta\mathbf{x}_i = \mathbf{F}_{i,i-1}[\mathbf{x}_{i-1} - \hat{\mathbf{x}}_{i-1}] + \mathbf{B}_{i-1}[\mathbf{u}_{i-1} - \mathbf{u}_{i-1}] + \mathbf{W}_{i-1}\mathbf{w}_{i-1}$. Because $\mathbf{u}_i$ is deterministic, the linearized Eq. (2.47) and (2.48) can be written as:

$$\mathbf{x}_i = \mathbf{f}_{i,i-1}(\hat{\mathbf{x}}_{i-1}, \mathbf{u}_{i-1}, \mathbf{0}) + \mathbf{F}_{i,i-1}[\mathbf{x}_{i-1} - \hat{\mathbf{x}}_{i-1}] + \mathbf{W}_{i-1}\mathbf{w}_{i-1} \tag{2.49}$$
$$\mathbf{y}_i = \mathbf{h}_i(\hat{\mathbf{x}}_i^-, \mathbf{0}) + \mathbf{H}_i[\mathbf{x}_i - \hat{\mathbf{x}}_i^-] + \mathbf{V}_i\mathbf{v}_i \tag{2.50}$$

The Jacobian matrices needed in the EKF are given by:

$$\mathbf{F}_{i,i-1} = \left[\begin{array}{ccc} \frac{\partial \mathrm{f}_1}{\partial x_1} & \cdots & \frac{\partial \mathrm{f}_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{f}_n}{\partial x_1} & \cdots & \frac{\partial \mathrm{f}_n}{\partial x_n} \end{array}\right] |(\hat{\mathbf{x}}_{i-1}, \mathbf{u}_{i-1}, \mathbf{0}) \tag{2.51}$$

$$\mathbf{W}_{i-1} = \left[\begin{array}{ccc} \frac{\partial \mathrm{f}_1}{\partial w_1} & \cdots & \frac{\partial \mathrm{f}_1}{\partial w_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{f}_n}{\partial w_1} & \cdots & \frac{\partial \mathrm{f}_n}{\partial w_r} \end{array}\right] |(\hat{\mathbf{x}}_{i-1}, \mathbf{u}_{i-1}, \mathbf{0}) \tag{2.52}$$

$$\mathbf{H}_i = \left[\begin{array}{ccc} \frac{\partial \mathrm{h}_1}{\partial x_1} & \cdots & \frac{\partial \mathrm{h}_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{h}_m}{\partial x_1} & \cdots & \frac{\partial \mathrm{h}_m}{\partial x_n} \end{array}\right] |(\hat{\mathbf{x}}_i^-, \mathbf{0}) \tag{2.53}$$

$$\mathbf{V}_i = \left[\begin{array}{ccc} \frac{\partial \mathrm{h}_1}{\partial v_1} & \cdots & \frac{\partial \mathrm{h}_1}{\partial v_s} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{h}_m}{\partial v_1} & \cdots & \frac{\partial \mathrm{h}_m}{\partial v_s} \end{array}\right] |(\hat{\mathbf{x}}_i^-, \mathbf{0}) \tag{2.54}$$

with $\mathbf{F}_{i,i-1}$ and $\mathbf{W}_{i-1}$ of dimensions $(n \times n)$ and $(n \times r)$, respectively and calculated at the a posteriori state estimate of the previous update. $\mathbf{H}_i$ and $\mathbf{V}_i$ are respectively of dimensions $(m \times n)$ and $(m \times s)$ and calculated at the a priori state estimate $\hat{\mathbf{x}}_i^- = \mathbf{f}_{i,i-1}(\hat{\mathbf{x}}_{i-1}, \mathbf{u}_{i-1}, \mathbf{0})$. The scalar functions f_n and h_m are as defined in section 2.3.1. For the linearized KF, the Jacobian matrices should be calculated in $(\hat{\mathbf{x}}_{i-1}^*, \mathbf{u}_{i-1}, \mathbf{0})$ for $\mathbf{F}_{i,i-1}$ and $\mathbf{W}_{i-1}$ and in $(\hat{\mathbf{x}}_i^{*,-}, \mathbf{0})$ for $\mathbf{H}_i$ and $\mathbf{V}_i$, with $\hat{\mathbf{x}}_i^{*,-}$ positioned on the nominal state-space trajectory.

The optimal estimate $\hat{\mathbf{x}}_i$ can be written as:

$$\boxed{\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \mathbf{K}_i[\mathbf{y}_i - \mathbf{h}_i(\hat{\mathbf{x}}_i^-, \mathbf{0})]} \tag{2.55}$$

with the a priori estimate given by the forward equation $\hat{\mathbf{x}}_i^- = \mathbf{f}_{i,i-1}(\hat{\mathbf{x}}_{i-1}, \mathbf{u}_{i-1}, \mathbf{0})$ and the update of the error covariance matrix given by:

$$\mathbf{P}_i^- = \mathbf{F}_{i,i-1}\mathbf{P}_{i-1}\mathbf{F}_{i,i-1}^T + \mathbf{W}_{i-1}\mathbf{Q}_{i-1}\mathbf{W}_{i-1}^T \tag{2.56}$$

The a posteriori estimation matrix is minimized for the optimal Kalman gain $\mathbf{K}_i$:

$$\boxed{\mathbf{K}_i = \mathbf{P}_i^-\mathbf{H}_i^T[\mathbf{H}_i\mathbf{P}_i^-\mathbf{H}_i^T + \mathbf{V}_i\mathbf{R}_i\mathbf{V}_i^T]^{-1}} \tag{2.57}$$

and with this $\mathbf{K}_i$ the a posteriori $\mathbf{P}_i$ can be calculated as:

$$\mathbf{P}_i = \mathbf{P}_i^- - \mathbf{K}_i\mathbf{H}_i\mathbf{P}_i^- \tag{2.58}$$

It is fundamental to understand that in the EKF the distributions (or densities) of the various random variables are no longer normal after undergoing a non-linear transformation. The EKF only approximates optimality by linearization. Julier

and Uhlmann [2004] worked on a variation of the EKF, trying to preserve normal distributions throughout the non-linear transformations.

Since earth processes are non-linear, many reported 'linear' applications of the KF, were EKF applications, e.g. by Hoeben and Troch [2000] and Walker *et al.* [2001b]. A review of KF applications in hydrology is given in chapter 7.

2.5.9 Ensemble Kalman filter (EnKF)

Both the KF and the EKF explicitly propagate the state error covariance matrices with an equation like Eq. (2.40) or (2.56). In case of non-linearities, these linearized equations are only an approximation. Furthermore, the integration of Eq. (2.40) is often hardly computionally feasible. Evensen [1994b] used an ensemble of model trajectories (Monte-Carlo approach, dynamical ensembles) to determine the error covariances directly from the spread of the states in an ensemble at a certain point in time, instead of obtaining a value for the error covariance matrix calculated with an approximate linearized equation. The model error is included in the ensemble perturbation and its covariance is not explicitly needed anymore for the propagation of the state error covariances.

Consider again the non-linear discrete system, given by:

$$\mathbf{x}_i = \mathbf{f}_{i,i-1}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1}, \mathbf{w}_{i-1}) \tag{2.59}$$

$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i, \mathbf{v}_i) \tag{2.60}$$

In contrast to the KF and the EKF, the model error $\mathbf{w}_i$ is assessed and explicitly added to the model equations, e.g. by allowing uncertain parameters to vary within a certain range and/or by perturbing the forcings. Simulation of model errors, including possible time correlated errors is discussed by Evensen [2003]. The a priori estimated (forecasted) ensemble member $\hat{\mathbf{x}}^-_{j,i}$ (with $j = 1, \ldots, N$ and N the number of members in the ensemble) is given by:

$$\hat{\mathbf{x}}^-_{j,i} = \mathbf{f}_{i,i-1}(\hat{\mathbf{x}}_{j,i-1}, \mathbf{u}_{i-1}, \mathbf{w}_{j,i-1}) \tag{2.61}$$

with $\mathbf{w}_{j,i}$ a random component drawn from a zero-mean Gaussian distribution with a predefined standard deviation. In the EnKF it is advised to explicitly account for the measurement errors to assure sufficient spread [Burgers *et al.*, 1998], by perturbing the measurements by a random realization of $\mathbf{v}_{j,i}$. Consequently, the measurement $\mathbf{y}_i$ is replaced by:

$$\mathbf{y}_{j,i} = \mathbf{y}_i + \mathbf{v}_{j,i} \tag{2.62}$$

Note the vectors $\mathbf{w}_{j,i}$ and $\mathbf{v}_{j,i}$ are realizations that vary for each ensemble member. Remark that for the EnKF, the model error is not assumed to be necessarily additive as for the (E)KF. Now the optimal estimate $\hat{\mathbf{x}}_{j,i}$ (analysis) of an ensemble member can be found as:

$$\boxed{\hat{\mathbf{x}}_{j,i} = \hat{\mathbf{x}}^-_{j,i} + \mathbf{K}_i[\mathbf{y}_{j,i} - \mathbf{h}_i(\hat{\mathbf{x}}^-_{j,i}, \mathbf{0})],\ \forall\ j = 1, \ldots, N} \tag{2.63}$$

The optimal gain $\mathbf{K}_i$ is identical for all members and just as in the KF and the EKF given by:

$$\boxed{\mathbf{K}_i = \mathbf{P}_i^- \mathbf{H}_i^T [\mathbf{H}_i \mathbf{P}_i^- \mathbf{H}_i^T + \mathbf{R}_i]^{-1}} \tag{2.64}$$

However, now $\mathbf{P}_i^-$ is defined slightly different as

$$\mathbf{P}_i^- = E[(\bar{\hat{\mathbf{x}}}_i^- - \hat{\mathbf{x}}_{j,i}^-)(\bar{\hat{\mathbf{x}}}_i^- - \hat{\mathbf{x}}_{j,i}^-)^T] \tag{2.65}$$

with $[\bar{\hat{\mathbf{x}}}_i^- - \hat{\mathbf{x}}_{j,i}^-]$ the a priori estimate error, and $\bar{\hat{\mathbf{x}}}_i^-$ the ensemble mean of the a priori estimates, given by:

$$\bar{\hat{\mathbf{x}}}_i^- = \frac{1}{N} \sum_{j=1}^{N} \hat{\mathbf{x}}_{j,i}^- = \hat{\mathbf{x}}_i^- \tag{2.66}$$

It is assumed that $\bar{\hat{\mathbf{x}}}_i^-$ is the best representation of the truth, which might sometimes end up as a weakness for the EnKF. The notation $\bar{\cdot}$ is introduced to highlight the ensemble average, but in essence, $\bar{\hat{\mathbf{x}}}_i^- = \hat{\mathbf{x}}_i^-$. In practice, $\mathbf{P}_i^-$ can be calculated as:

$$\mathbf{P}_i^- = \frac{1}{N-1} \left[[\bar{\hat{\mathbf{x}}}_i \cdots \bar{\hat{\mathbf{x}}}_i] - [\hat{\mathbf{x}}_{1,i}^- \cdots \hat{\mathbf{x}}_{N,i}^-] \right] \left[[\bar{\hat{\mathbf{x}}}_i \cdots \bar{\hat{\mathbf{x}}}_i] - [\hat{\mathbf{x}}_{1,i}^- \cdots \hat{\mathbf{x}}_{N,i}^-] \right]^T \tag{2.67}$$

Analogously, $\mathbf{P}_i$ is calculated as $\mathbf{P}_i = E[(\bar{\hat{\mathbf{x}}}_i - \hat{\mathbf{x}}_{j,i})(\bar{\hat{\mathbf{x}}}_i - \hat{\mathbf{x}}_{j,i})^T]$ with $[\bar{\hat{\mathbf{x}}}_i - \hat{\mathbf{x}}_{j,i}]$ the a posteriori (analysis) estimate error, and $\bar{\hat{\mathbf{x}}}_i$ the ensemble mean of the a posterior estimates. Furthermore, the error covariance of the analyzed ensemble can be reduced in the same way as in the (extended) KF:

$$\mathbf{P}_i = \mathbf{P}_i^- - \mathbf{K}_i \mathbf{H}_i \mathbf{P}_i^- \tag{2.68}$$

This matrix is not needed in the filter calculation and only serves as a measure for the accuracy of the a posteriori estimate. Note that in the KF and the EKF, the estimation error is defined as the deviation of the estimation from the real state. However, as this real state is not known, in the EnKF the real state is approximated by the ensemble mean.

With $\mathbf{v}_{j,i}$ a realization of a perturbation on the observation $\mathbf{y}_i$, the covariance matrix $\mathbf{R}_i$ can be found as:

$$\mathbf{R}_i = \frac{1}{N-1} [\mathbf{v}_{1,i} \;\; \mathbf{v}_{2,i} \;\; \cdots \;\; \mathbf{v}_{N,i}] \, [\mathbf{v}_{1,i} \;\; \mathbf{v}_{2,i} \;\; \cdots \;\; \mathbf{v}_{N,i}]^T \tag{2.69}$$

If a matrix is defined with as columns all ensemble members of the estimated state $\hat{\mathbf{X}}_i = [\hat{\mathbf{x}}_{1,i} \;\; \ldots \;\; \hat{\mathbf{x}}_{N,i}]$ and a matrix with columns of all perturbed realizations of one measurement vector $\mathbf{Y}_i = [\mathbf{y}_{1,i} \;\; \ldots \;\; \mathbf{y}_{N,i}]$, then the analysis equation for the a posteriori optimal estimate can be written as:

$$\hat{\mathbf{X}}_i = \hat{\mathbf{X}}_i^- + \mathbf{K}_i [\mathbf{Y}_i - \mathbf{H}_i \hat{\mathbf{X}}_i^-] \tag{2.70}$$

Further practical formulations are given by Evensen [2003]. A presentation of the EnKF using representers (influence functions for each assimilated observation) is given by Evensen and van Leeuwen [2000]. Practical applications will be discussed in chapter 7 and further.

Data selection

Houtekamer and Mitchell [1998] performed a data selection within an EnKF, which should typically not be done with a KF-based algorithm, but which is similar to statistical interpolation techniques. They proposed a cutoff radius for the correlation, which was a horizontal distance beyond which no data were assimilated. This way observations were eliminated that were weakly correlated with the analysis point and the error covariance matrices were greatly reduced. They indicated that very large ensemble sizes would be needed to estimate weak correlations accurately. However, imposing a cutoff radius resulted in imbalances in the analysis and therefore, Houtekamer and Mitchell [2001] used another method, which did not necessarily always remove balance problems [Mitchell *et al.*, 2002; Lorenc, 2003], but which was reported to be superior. Because the finite ensemble size caused estimated correlations to be noisy, a Hadamard or Schur (elementwise) product of the covariances of the background obtained from the ensembles and a correlation function with local support [Gaspari and Cohn, 1999] was calculated. More specifically, the Kalman gain is calculated then by :

$$\mathbf{K}_i = \rho \bullet \mathbf{P}_i^- \mathbf{H}_i^T [\mathbf{H}_i \rho \bullet \mathbf{P}_i^- \mathbf{H}_i^T + \mathbf{R}_i] \tag{2.71}$$

with $\rho \bullet \mathbf{P}_i^-$ the Schur product of a correlation matrix $\mathbf{A}$ with a covariance matrix $\mathbf{P}_i^-$, with each resulting element $C_{\mathrm{x,y}}$ given by $A_{\mathrm{x,y}} P_{\mathrm{x,y}}^-$, with $P_{\mathrm{x,y}}^-$ an element of matrix $\mathbf{P}_i^-$. The element $A_{\mathrm{x,y}}$ is obtained from the correlation function ρ for the distance between points x and y. This covariance localization is also used to increase the rank of the experimentally determined $\mathbf{P}_i^-$ (see later).

2.5.10 Variations to the Kalman filter

Brasseur [2004] discussed several alternatives to the standard linear KF equations. These include simplified schemes, such as optimal interpolation in which $\mathbf{P}_i^-$ is predefined and mostly fixed in time. Another simplification can be introduced by solving the (Ricatti) equation for $\mathbf{P}^-$ through iteration prior to the assimilation sequence, since $\mathbf{P}^-$ converges to a steady-state solution, given a time-invariable $\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$ and $\mathbf{R}$ [Fukumori *et al.*, 1993].

Order reduction is an alternative way to make the KF practical for use in large sized problems. The dimensionality can be reduced in the state space [Dee, 1991; Cane *et al.*, 1996], where statistical properties of the state are transformed to a low-dimension space. A reduction can also be performed in the error space [Cohn and Todling, 1996; Lermusiaux, 1999; Verlaan and Heemink, 1997; Pham *et al.*, 1998], with the intention to correct the model state along the most representative directions of the forecast. Low-rank error covariance matrices have been proposed to simplify the computations.

Apart from the EKF and EnKF, other extensions to the KF have been developed for non-linear systems. Better approximations of the non-linear transformations

have been proposed [Schei, 1997; Nørgaard *et al.*, 2000] to improve the capability of the EKF. Unscented filtering [Julier and Uhlmann, 2004; Lefebvre *et al.*, 2002] is similar to EnKF, but instead of completely random perturbation through Monte Carlo generation, a set of appropriately chosen weighted points are used to parameterize the means and covariances of the probability distributions. Another method which is also well suited for non-linear applications is the particle filter, which is a sequential Monte Carlo method [Doucet *et al.*, 2001].

2.6 Variational assimilation

Variational assimilation is another approach to solve the same state estimation problem. With this method, the optimal state estimate is chosen as the state vector that minimizes a cost function. As opposed to the method of the Kalman filter, variational algorithms do not provide directly an estimate of the analysis error.

Cost functional $\mathcal{J}$

Variational data assimilation is formulated as an optimization problem, based on optimal control theory and can be solved using classical numerical methods, rather than stochastic calculus in estimation theory. The optimal state, which is a model solution, if the model is assumed to be perfect (strong constraint), is found in variational data assimilation by minimizing a cost function that penalizes the distance between observations and model results. The cost or penalty function $\mathcal{J}$ is a functional of the state vector $\mathbf{x}$ (control variable), that penalizes the misfit between the guess (current analysis) of the true state and two possible representations of the true state: observations and model forecasts. It is possible to consider the distances or misfits over a time interval or at one point in time, resulting in respectively a smoothing or filtering problem. When dealing with linear models, the cost function is typically quadratic. However, it should be noticed that non-linear models lead to non-quadratic cost functions, which -due to the shape- may result in complications in finding the optimal state.

There are 2 standard formulations of the minimization problem: strong and weak constraint formulations. In case the model is assumed to be 'perfect', the optimal state to be found will be an exact solution of the model dynamics: a strong constraint [Le Dimet and Talagrand, 1986] is imposed to the inverse problem. This leads to the most compact formulation of the cost function. Possible additional constraints can be added in the cost function, taking into account e.g. model error. Expansion of classic inverse problems to the weak constraint fit [Sasaki, 1970] of both model dynamics and observations leads to generalized inverse problems [Bennett, 1992]. In the weak constraint formulation, prespecified weights are often used in the cost functional.

Minimization techniques

There are basically 2 ways to find the optimal state $\hat{\mathbf{x}}$ in variational assimilation. Direct minimization of the cost function (e.g., Liang [2004]) uses mostly iterative descent algorithms. This involves evaluations of the cost function and its gradient and iterative updates of the state $\mathbf{x}_i$ at a given instant or $\mathbf{x}_0$ at the beginning of the variational assimilation window for smoothing (4D-Var, see further). Usually, $\mathbf{x}$ is updated by adding a correction proportional to $-\nabla\mathcal{J}(\mathbf{x})$, which is the gradient of the cost function with a magnitude that corresponds to the vector form of the usual derivative, as in the steepest descent method. A more efficient but similar descent method is the conjugate gradient method. Other direct minimization techniques are Newton and quasi-Newton methods, simulated annealing and genetic algorithms. Simulated annealing is especially well suited for strongly non-linear inverse problems, where methods like gradient descent tend to fail [Evensen, 1994a]. Adjoint techniques, as used in 4D-Var (strong constraint, see further), offer an efficient method to solve the gradient $\nabla\mathcal{J}(\mathbf{x})$ and can be readily derived using calculus of variations. Courtier *et al.* [1993] compiled a list of literature references addressing mainly applications of adjoint equations to meteorological problems.

Another approach to the minimization problem is to obtain extrema by setting the variation δ of the cost function, $\delta\mathcal{J}$, to 0, leading to Euler-Lagrange equations. In case of a weak constraint problem, model error typically couples the state evolution in the Euler-Lagrange equations with the adjoint evolution. The representer technique [Reichle, 2000] is an optimal method to solve this kind of problem in case the model is linear. For non-linear dynamics, an iteration technique with linearization is used.

2.6.1 Three-dimensional variational assimilation (3D-Var)

In 3D-Var, the 3D refers to the 3 dimensions in space that may be covered by the system state. For discrete (non-)linear systems

$$\mathbf{x}_i = \mathbf{f}_{i,i-1}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1}) + \mathbf{w}_i \tag{2.72}$$

$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i) + \mathbf{v}_i \tag{2.73}$$

the variational statistical formulation for a multi-dimensional (vector) estimation problem can be expressed by the cost function $\mathcal{J}$ at a certain time step i:

$$\mathcal{J}(\mathbf{x}_i) = \frac{1}{2}[\mathbf{x}_i - \hat{\mathbf{x}}_i^-]^T \mathbf{P}_i^{-^{-1}} [\mathbf{x}_i - \hat{\mathbf{x}}_i^-] + \frac{1}{2}\left[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)\right]^T \mathbf{R}_i^{-1}\left[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)\right] \tag{2.74}$$

The state vector $\mathbf{x}_i$ is the control vector which will minimize the cost function and yields the a posteriori estimate $\hat{\mathbf{x}}_i$, that best fits the background information $\hat{\mathbf{x}}_i^-$ as well as the observations, given their respective degree of accuracy $\mathbf{P}_i^{-^{-1}}$ and $\mathbf{R}_i^{-1}$. For ease in computation $\mathbf{f}_{i,i-1}$ and $\mathbf{h}_i$ are usually linearized to the operators

$\mathbf{F}_{i,i-1}$ and $\mathbf{H}_i$. Setting the derivative of $\mathcal{J}(\mathbf{x}_i)$ to null, yields:

$$0 = \mathbf{P}_i^{-^{-1}}[\mathbf{x}_i - \hat{\mathbf{x}}_i^-] - \mathbf{H}_i^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i] \tag{2.75}$$

so that $\mathbf{x}_i$ can be estimated by:

$$\hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^- + \left[\mathbf{P}_i^{-^{-1}} + \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i\right]^{-1} \mathbf{H}_i^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{H}_i \hat{\mathbf{x}}_i^-] \tag{2.76}$$

This solution is similar to the KF solution, with the important difference that the variational solution does not give any information on the error on the estimate.

Instead of solving Eq. (2.76), the solution can be sought iteratively by performing several evaluations of the cost function and the gradient to approach the minimum using e.g. a suitable descent algorithm, starting from an initial guess $\hat{\mathbf{x}}_i^g$, which is mostly chosen, but does not need, to be $\hat{\mathbf{x}}_i^-$.

2.6.2 Four-dimensional variational assimilation (4D-Var)

4D-Var is an extension of 3D-Var, by including a time dimension. It is sometimes referred to as a dynamic observer assimilation method and it can be considered as an optimization problem in which the state at the beginning of the assimilation window is to be found by minimizing the a priori residuals (observation-minus-forecasts) over that time window. Consider a discrete (non)-linear system given by:

$$\mathbf{x}_i = \mathbf{f}_{i,i-1}(\mathbf{x}_{i-1}, \mathbf{u}_{i-1}) \tag{2.77}$$

$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i) + \mathbf{v}_i \tag{2.78}$$

assuming a 'perfect' model (strong constraint), with $\mathbf{w}_i = 0$. With this assumption, the cost function over an interval of T time steps is given by:

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}[\mathbf{x}_0 - \hat{\mathbf{x}}_0^-]^T \mathbf{P}_0^{-^{-1}}[\mathbf{x}_0 - \hat{\mathbf{x}}_0^-] + \frac{1}{2}\sum_{i=0}^{\mathrm{T}-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)] \tag{2.79}$$

with $\mathbf{x}_0$ (and $\mathbf{x}_i$) the vector which is to be optimized. The first term is called the background penalty $\mathcal{J}_b$, and the second term is the observation penalty $\mathcal{J}_o$. Finding the optimal $\hat{\mathbf{x}}_0$ for this cost functional is clearly a non-linear constrained optimization problem, which is quite difficult to solve in general. Therefore, a first simplification is that the operators $\mathbf{f}_{i,i-1}()$ and $\mathbf{h}_i()$ will be linearized and in further computations will be replaced by their tangent linear models $\mathbf{F}_{i,i-1}$ and $\mathbf{H}_i$. A sufficient condition for $\mathcal{J}(\mathbf{x}_0)$ to have only one minimum is that its shape is strictly convex. This can only be assured if $\mathcal{J}$ is quadratic, i.e. when the dynamics and observations are represented by linear functions. Secondly, the forward (forecast) model will be expressed as the product of intermediate forecast steps (causality, i.e. a state at time step i does not depend on input or states at time steps $> i$),

such that $\mathbf{x}_i$ is written as the result of a model integration with initial condition $\mathbf{x}_0$:

$$\mathbf{x}_i = \mathbf{f}_{i,i-1} \circ \mathbf{f}_{i-1,i-2} \cdots \circ \mathbf{f}_{1,0}(\mathbf{x}_0) \quad \text{or} \quad \mathbf{x}_i = \mathbf{F}_{i,i-1}\mathbf{F}_{i-1,i-2} \ldots \mathbf{F}_{1,0}\mathbf{x}_0 \tag{2.80}$$

with $\circ$ referring to the 'after'-operator. In the quadratic functional $\mathcal{J}$, one assumes Gaussian errors. If this assumption is biased, the penalty function defines no longer a maximum-likelihood estimator, but still a variance-minimizing estimator.

The major operator in the residual $[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]$ is the observational forward model $\mathbf{h}_i()$, acting on the model state $\mathbf{x}_i$. The forward model produces this model state from the guess of the analyzed state $\mathbf{x}_0$, so:

$$\mathbf{h}_i(\mathbf{x}_i) = \mathbf{h}_i(\mathbf{f}_{i,i-1} \circ \mathbf{f}_{i-1,i-2} \cdots \circ \mathbf{f}_{1,0}(\mathbf{x}_0)) \tag{2.81}$$

Finding the best guess for $\mathbf{x}_0$, i.e. in this case the initial conditions, requires finding the elements of $\hat{\mathbf{x}}_0$ which minimize the value of $\mathcal{J}$. For the most common minimization methods, the gradient of $\mathcal{J}$, rather than the value of $\mathcal{J}$ itself is needed. The algorithm requires an initial guess $\hat{\mathbf{x}}^g$ for the a posteriori state estimate. In many implementations the first guess is simply assigned the value of the background or a priori state estimate $\hat{\mathbf{x}}_i^-$. The gradient with respect to the initial conditions is given by:

$$\nabla_{\mathbf{x}_0}\mathcal{J} = \left[\frac{\partial \mathcal{J}}{\partial \mathbf{x}_0}\right]^T = \mathbf{P}_0^{-1}[\mathbf{x}_0 - \hat{\mathbf{x}}_0^-] - \sum_{i=0}^{\mathrm{T}-1} \left[\frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_0}\right]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)] \tag{2.82}$$

using the feature of symmetric covariance matrices. Note that $\mathbf{x}_i$ represents a time series, while $\mathbf{x}_0$ is only a vector at one (initial) time, meaning that the term $\left[\frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_0}\right]^T$ gives the sensitivity of a certain element of the model-predicted observation vector at time step i due to changes in an element of the initial condition vector. A more appealing Jacobean for $\mathbf{h}_i$ would be $\left[\frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_i}\right]^T$. Recalling the relationship between $\mathbf{x}_i$ and $\mathbf{x}_0$, using the tangent-linear model, the chain rule for derivatives gives:

$$\left[\frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_0}\right]^T = \left[\frac{d\mathbf{x}_i}{d\mathbf{x}_0}\right]^T \left[\frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_i}\right]^T = \left(\mathbf{F}_{i,i-1}\mathbf{F}_{i-1,i-2} \ldots \mathbf{F}_{1,0}\right)^T \left[\frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_i}\right]^T \tag{2.83}$$

The operator $\left(\mathbf{F}_{i,i-1}\mathbf{F}_{i-1,i-2} \ldots \mathbf{F}_{1,0}\right)^T$ is called the adjoint operator, acting directly on the adjoint variable $\left[\frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_i}\right]^T$. While $\mathbf{F}_{i,0}$ transmits information from $\mathbf{x}_0$ to $\mathbf{x}_i$, the adjoint operator $\mathbf{F}_{i,0}^T$ transmits information in the reverse direction, but for the adjoint variables.

Inserting this transformation and linearizing the operators (for practical simplicity) yields:

$$\nabla_{\mathbf{x}_0}\mathcal{J} \simeq \mathbf{P}_0^{-1}[\mathbf{x}_0 - \hat{\mathbf{x}}_0^-] - \sum_{i=0}^{\mathrm{T}-1} \mathbf{F}_{i,0}^T \mathbf{H}_i^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)] \tag{2.84}$$

with the linearized (around the background $\hat{\mathbf{x}}_0^-$) adjoint time evolution operator:

$$\mathbf{F}_{i,0}^T = [\mathbf{F}_{i,i-1}\mathbf{F}_{i-1,i-2}\cdots\mathbf{F}_{1,0}]^T \tag{2.85}$$

and the effective linearization of the forward observation operator:

$$\mathbf{H}_i = \frac{d\mathbf{h}_i(\mathbf{x}_i)}{d\mathbf{x}_i} \tag{2.86}$$

$\mathbf{F}_{i+1,i}^T$ is usually called the adjoint of the linearized model and $\mathbf{H}_i^T$ the adjoint of the linearized observation operator. In this case the adjoint is identical to the transpose, therefore the common adjoint notation $(.)^*$ is replaced by the transpose symbol $(.)^T$.

The evaluation of $\nabla_{\mathbf{x}_0}\mathcal{J}$ can be performed without any of the above derivations, but just by calculating finite differences. Another method is direct evaluation of Eq. (2.84), involving calculation of the second term for every time step. It is clear that this involves e.g. T times the execution of $\mathbf{F}_{1,0}$, T − 1 times the execution of $\mathbf{F}_{2,1}$, etc. The adjoint technique calculates the second term more efficiently, requiring only one adjoint integration backwards over the whole time interval, i.e.:

$$\nabla_{\mathbf{x}_0}\mathcal{J}_o \simeq \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{r}_0 + \mathbf{F}_{1,0}^T\left[\mathbf{H}_1^T\mathbf{R}_1^{-1}\mathbf{r}_1 + \mathbf{F}_{2,1}^T[\ldots + \mathbf{F}_{\mathrm{T}-1,\mathrm{T}-2}^T\mathbf{H}_{\mathrm{T}-1}^T\mathbf{R}_{\mathrm{T}-1}^{-1}\mathbf{r}_{\mathrm{T}-1}]\right] \tag{2.87}$$

with $\mathbf{r}_i = [\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)]$. For a comprehensive explanation of this method one is referred to Bouttier and Courtier [1999] and Bannister [2001]. Summarized, the adjoint technique consists in 2 main actions: (i) the forward integration of the model states by the forward time evolution model to obtain $\hat{\mathbf{x}}_i$ and their residuals $\mathbf{r}_i$ and (ii) the backward integration of the adjoint variables by the adjoint model. This will yield the gradient of the observation penalty $\mathcal{J}_o$ or the second term of Eq. (2.84). The solution is $\hat{\mathbf{x}}_0$ for which the minimal $\mathcal{J}$ is found. The development of adjoints is mostly a difficult task for the non-linear hydrological models, and impractical to derive for existing models, even though adjoint compilers are available. Some practical examples are discussed by Reichle *et al.* [2001b] and Castelli *et al.* [1999].

2.6.3 Alternatives to 3D/4D-Var

Instead of estimating the complete state, Courtier *et al.* [1994] proposed the incremental 3D/4D-Var in which the analysis increment only is estimated. Often a linear simplification operator is introduced in order to reduce the dimension of the minimization problem and the related computational cost by making the control variable smaller than the model state. In case the optimal state is sought by direct minimization, the minimization space is often changed to improve the conditioning of the problem. This technique is called preconditioning (e.g. Tziperman *et al.* [1992]). The condition number of a cost function is a quantity that measures the

ellipticity of its iso-surfaces. The larger this number, the more ill-conditioned the problem is.

Instead of minimizing the cost function in the state space, Physical Space Assimilation Systems (PSAS) apply the minimization in the observation space, taking advantage of the smaller dimension of the observation vector, as described by Courtier [1997] and Cohn *et al.* [1998]. This dual formulation to variational assimilation was originally applied for spectral models. Where in 4D-Var the forward model is run first and the adjoint model afterwards, in 4D-PSAS the adjoint is run first. The PSAS algorithm is equivalent to the representer method [Reichle, 2000; Evensen, 1994b; Bennett, 1992].

2.6.4 Weak constraint variational data assimilation

Including a weak constraint leads to generalized inverse problems. There are two common approaches to the minimization problem. The first is setting $\delta\mathcal{J} = 0$, which mostly results in a system of coupled Euler-Lagrange equations. The coupling, which only occurs with inclusion of the weak constraint, is generally broken through application of the representer method [Reichle, 2000; Evensen, 1994a]. The second method is a direct minimization by an iterative numerical algorithm.

2.6.5 Kalman filter versus variational data assimilation

Given a finite time interval for modeling, and assuming a perfect model, the 4D-Var method and the Kalman filter yield the same result at the last moment of assimilation. Inside the time interval, 4D-Var is more optimal, because it uses all observations at once (before and after the time step of analysis), i.e. it is a smoother. A disadvantage of sequential methods is the discontinuity in the corrections, which causes model chocks. Through variational methods, there is a larger potential for dynamically based balanced analyses, which will always be situated within the model climatology. Operational 4D-Var assumes perfect models: no model error can be included. With inclusion of model error, coupled equations are to be solved for minimization. Through Kalman filtering it is in general simpler to account for model error.

Both the Kalman filter and 3D/4D-Var rely on the validity of the linearity assumption. Adjoints depend on this assumption and incremental 4D-Var is even more sensitive to linearity. Uncertainty estimates via the Hessian are critically dependent on a valid linearization. Further, with variational assimilation it is more difficult to obtain an estimate of the quality of the analysis or of the state's uncertainty after updating.

In the framework of estimation theory, the goal of variational assimilation is the estimation of the conditional mode (maximum a posteriori probability) estimate,

while for the Kalman filter the conditional mean (minimum variance) estimate is sought.

Hybrid assimilation methods have been explored, in which a sequential method is used to produce the a priori state error or background error covariance for variational assimilation.

2.7 Summary

For realistic system representations, the deterministic state space system should be replaced by a stochastic system representation to account for model noise. For a stochastic system, the state cannot simply be found as a solution of the system, but it is described by a pdf. The moments of the pdf are propagated in time and for Gaussian state distributions passed through linear systems, the first two moments suffice for a complete description. For non-linear systems the complete pdf should be calculated. If observations are available, estimation of the state can benefit from this source of information. However, observations are never perfect either and also contain a stochastic component. Therefore, in the process of state estimation, the uncertainty of both information sources is weighted in order to obtain a best possible estimate of the true state. The most common data assimilation techniques for state estimation in earth sciences are reviewed in this chapter, with special attention to the Kalman filter.

Part I

Data and Model

Chapter 3

OPE^3 data

3.1 Introduction

Since soil moisture is an important state variable of the land surface system, the collection and analysis of soil moisture data at different spatial and temporal scales received a lot of attention. At the regional scale, soil moisture is known to interact with the atmosphere, which is of major importance for weather predictions [Zhang and Frederiksen, 2003; Koster *et al.*, 2004], to influence the climate [Dirmeyer, 2000] and its change [Manabe and Delworth, 1990] and to have a controlling function in the hydrological cycle in general. At field scales, soil moisture has an impact on the generation of runoff [Kitanidis and Bras, 1980; Pauwels *et al.*, 2002] and erosion [Moore *et al.*, 1988], plant growth [Pauwels *et al.*, 2006] and the chemical behavior of fertilizers, which is important to agriculture and environment. Soil moisture itself is influenced by a combination of atmospheric forcings, terrain features, texture and vegetation.

At the field scale, soil moisture values are generally obtained through ground measurements [Walker, 1999; Walker *et al.*, 2004b], which are typically point measurements collected in 'representative' locations and at specific time instants. Detailed analysis of soil moisture patterns would require a dense network of observations, which is often impractical. Therefore, remote sensing offers a useful alternative to measuring soil moisture [Owe *et al.*, 1999; Vinnikov *et al.*, 1999; Walker *et al.*, 2004a]. However, through these techniques, only the top layer soil moisture is captured and the current data rarely provide any information on detailed field variability. Furthermore, remote sensing data need to be calibrated by ground truth data. Therefore, the collection and use of ground measurements remains a necessity.

Projects for the collection of new datasets of soil moisture have been initiated by the need to fully identify the land surface system [Georgakakos and Baumer, 1996].

It is through observed soil moisture patterns in space and time that natural processes can be understood and converted into physical laws, empirical relationships and model structures. Once a model structure is set up, observations are needed for parameterization. Since the final model will never be perfect, observations are useful to update the state variables and parameters, if the simulations tend to deviate from the truth.

In this chapter, field scale spatial and temporal characteristics of soil moisture and their importance for modeling purposes are discussed. Several authors [Western and Grayson, 1998; Petrone *et al.*, 2004; Teuling and Troch, 2005] discussed soil moisture characteristics at the field scale, but only a few datasets consist of measurements taken at a high temporal resolution and covering three dimensions in space, i.e. with dense measurements both horizontally over a given study area and vertically over a profile.

3.2 Optimizing Production Inputs for Economic and Environmental Enhancement (OPE^3) project

OPE^3 is an interdisciplinary research project which was started in 1998. This project addresses major environmental and economic issues facing U.S. agriculture. Researchers from different disciplines participate in the OPE^3 project, amongst others the Beltsville Agricultural Research Center (BARC) - Agricultural Research Service (ARS) of the U.S. Department of Agriculture (USDA) and the Hydrological Sciences Branch (HSB) of the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC).

The main objectives of the OPE^3 project [Gish *et al.*, 2000] are (i) the determination of atmospheric, surface and subsurface watershed-scale fluxes of water and agricultural chemicals from three crop production systems, (ii) the determination of the behavior and environmental impact of chemicals on a wooded riparian wetland and a first-order stream, (iii) the development of remotely sensed data products and analytical techniques for measuring and managing the spatial variability of crops and soils and (iv) the determination of long term economic and environmental impacts of these crop production systems and the evaluation of their trade-offs using integrated economic and biophysical simulation models. For more information one is referred to http://hydrolab.arsusda.gov/ope3/. Data from the OPE^3 field have been used in several studies on different research topics, such as chemical transit times and pesticide volatilization [Prueger *et al.*, 2005; Kung *et al.*, 2006]. In this work, data used from the OPE^3 project are mainly soil moisture and atmospheric forcing data. Because of the high spatial and temporal resolution of the data and its sufficient time length for modeling, the OPE^3 field provides an excellent opportunity to study soil moisture in detail. At the time of writing no detailed soil moisture profile simulations for the OPE^3 field were reported.

Chinkuyu *et al.* [2004] (and Chinkuyu *et al.* [2005]) studied how well hydrology processes were simulated by 2 agricultural water quality models. They aggregated soil moisture observations to daily values, considered only subfield averaged soil moisture and averaged the soil moisture over the upper 120 cm, which was taken as the rooting depth. As will be shown below, the variability over this depth is considerable and the rooting depth of corn is in reality only limited, so that no specific details of the hydrology could be obtained from this study. Furthermore, there were some inconsistencies in the observed data reported by Chinkuyu *et al.* [2004] and Chinkuyu *et al.* [2005].

In this study, some additional data from a nearby Soil Climate Analysis Network (SCAN) site were also used. These data are available through the World Wide Web (WWW) at http://www.wcc.nrcs.usda.gov/scan/ and additional information was made available by the National Water and Climate Center of the USDA Natural Resources Conservation Service (NRCS). The NRCS is a leader of a cooperative U.S. nationwide comprehensive soil moisture and climate information system. Through the SCAN, the NRCS integrates information from existing data networks and establishes new data collection points in collaboration with federal, state, local and tribal entities. The SCAN focuses on agricultural areas of the U.S..

In general, programs for on-site measurement of soil moisture are very useful for a broad range of modeling studies and are currently receiving particular attention for verification of remotely sensed soil moisture estimates. Some earlier programs in the U.S. were reviewed by Georgakakos and Baumer [1996].

3.3 Site description

3.3.1 Geographic characterization

The OPE^3 project is conducted on 4 sub-watersheds situated in Prince Georges County, Maryland, USA (figure 3.1). The site is located in the Anacostia watershed, which is part of the Middle Potomac-Anacostia-Occoquan watershed, cataloged by the U.S. Geological Survey (USGS) under Hydrologic Unit Code 02070010. Water draining from the field feeds a wooded riparian wetland and first-order stream, Beaver Dam Creek (see later, figure 3.7), which subsequently drains into the Anacostia river, the Potomac river and the Chesapeake Bay. Figure 3.2 shows the geographical position of the OPE^3 field site in the watershed.

Four adjacent watersheds with similar surface and subsurface soil and water flow characteristics were delineated in a 21 ha study area. The 4 hydrologically bounded sub-watersheds are about 4 ha each and surrounded by earthen berms. These watersheds are named A, B, C and D from North to South. The coordinates of the center of the B field are 39.031 N and 76.844 W. A Digital Elevation Model (DEM) was constructed by interpolation of data acquired using a dual frequency, carrier phase real-time kinematic global positioning system (GPS) with vertical

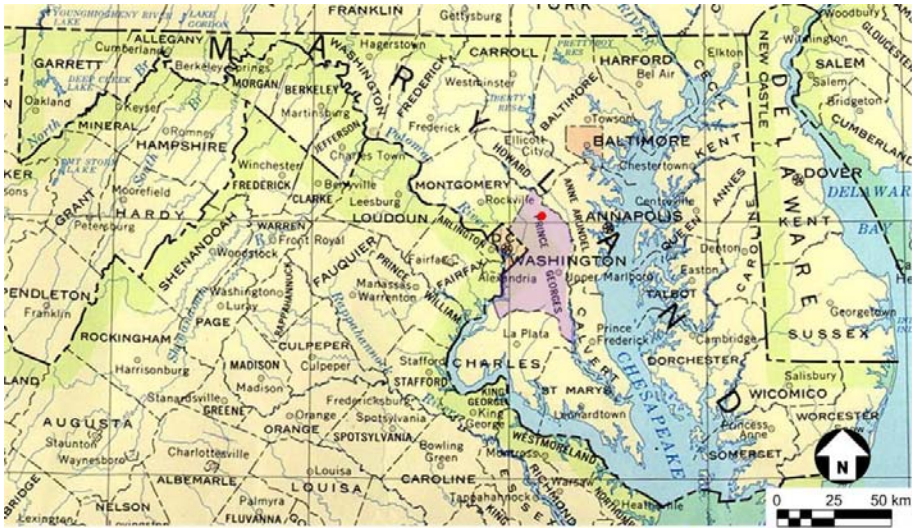


Figure 3.1: Indicative geographic position of the OPE^3 field site (red dot), adapted after USGS (1972, limited update 1990).

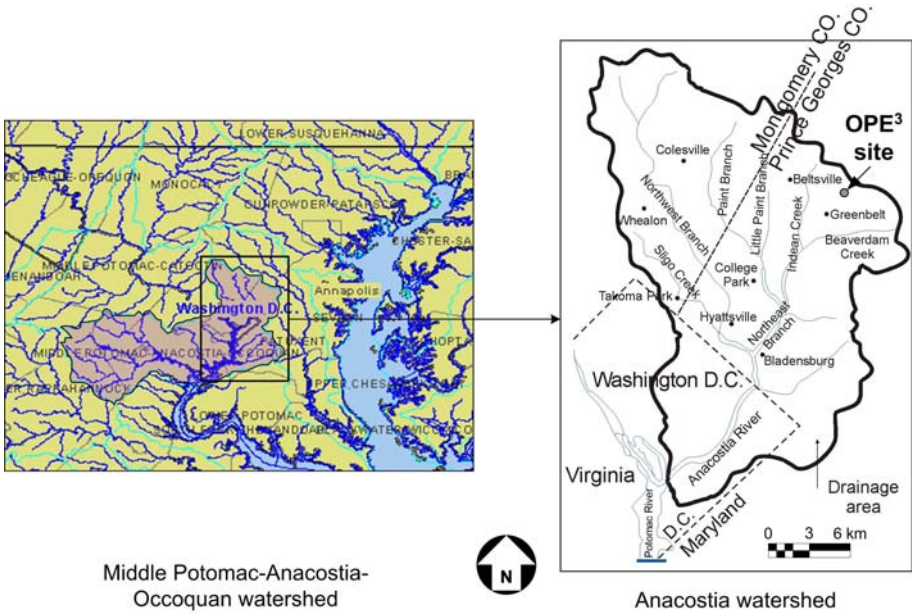


Figure 3.2: Drainage area of the Anacostia watershed with indication of the OPE^3 field site, adapted after the United States Environmental Protection Agency's (EPA) Enviromapper and the Anacostia Watershed Society.

accuracy within 5 cm [Dulaney *et al.*, 2000]. The data were recorded with irregular spaces ranging from less than a meter to maximal values of approximately 10 m. Figure 3.3 shows the 10 m resolution DEM, constructed by spatial interpolation with ordinary kriging [Goovaerts, 1997]. A Gaussian shaped variogram model with nugget effect,

$$\gamma(h) = C(0) + C_l \left[1 - \exp\left(\frac{-3h^2}{a^2}\right)\right] \tag{3.1}$$

was fitted to the experimental omni-directional variogram, with a nugget $C(0)$ of 0.02 m^2, a range a of 318.5 m and a sill C_l of 7.23 m^2, as shown in figure 3.3. The variogram values are function of the lag h between points in space. More information on variograms and their parameters will be provided in section 3.8.2. The average height is about 40 m above sea level and the slope is varying from 1 to 4 %.

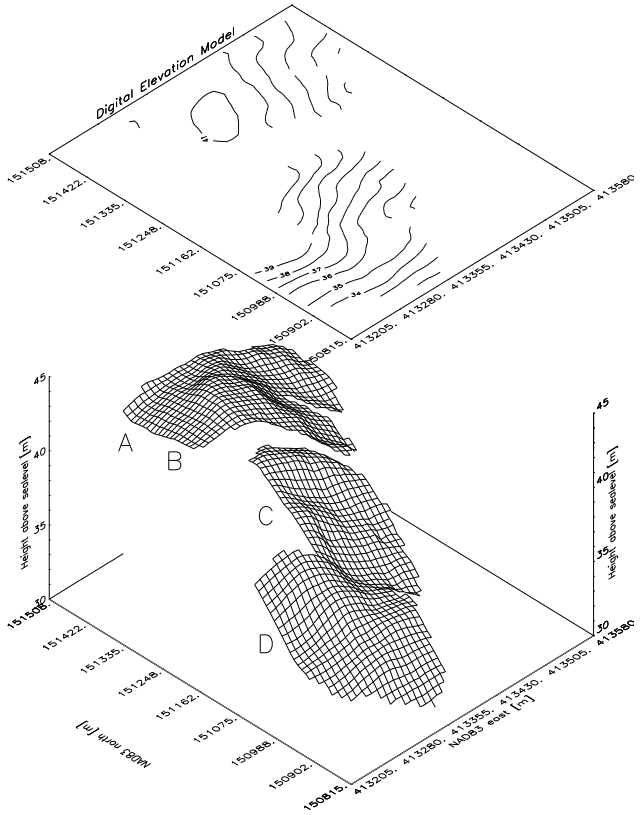
3.3.2 Geologic characterization

The geologic map of Prince Georges County in figure 3.4, based on the Geologic map of Maryland produced by the Maryland Geological Survey (1968), shows that the major geological formation in the area of the research site dates from the Cretaceous, with sand, gravel and clay as main mineral products.

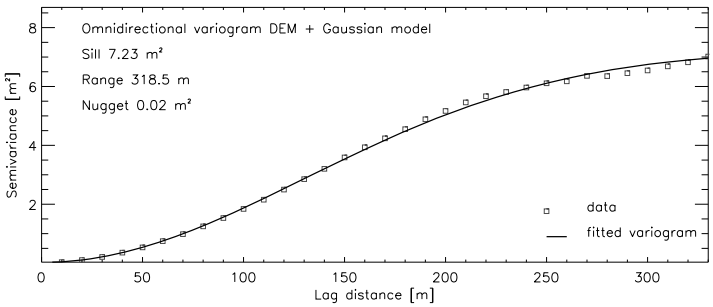
Each of the sub-watersheds of the OPE^3 field was formed from sandy fluvial deposits. In figure 3.5, the percentages clay, sand and silt are shown for the upper 9 cm (sampled in 1999). The measured data were spatially interpolated by ordinary kriging, using spherical variograms for the 3 texture classes. The spherical variogram is given by:

$$\gamma(h) = C(0) + C_l \left[\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right] \tag{3.2}$$

The nuggets $C(0)$ are 1.13, 7.00 and 10.00 %^{2} for clay, silt and sand respectively, the sills C_l 5.04 %^{2}, 26.0 %^{2} and 49.0 %^{2} and the ranges a 300 m, 205 m and 218 m, respectively. More information on variograms and their parameters is again given in section 3.8.2. Since the ranges for the different texture classes are different, one should be careful to obtain a sum of 100 % when the different classes are interpolated independently. However, because of the high data density, this concern is of limited importance. Based on the interpolated data the soil can be described as sandy loam according to the USGS soil classification, with an average of (15.62 ± 1.63) % clay, (22.19 ± 4.07) % silt and (62.17 ± 5.56) % sand. A clay layer is present under the entire site, varying from 0.9 to 3.5 m below the soil surface [Gish *et al.*, 2002]. Due to the coarse sand with abundant gravel, the water holding capacity of the deeper layers is very low (10 vol% or less), and the water content only increases due to a rising perched water table or due to preferential flow. Gish *et al.* [2002] already discussed that the most important type of preferential flow observed in this site is funnel flow which is gravity driven.



(a) Digital Elevation Model at 10 m resolution.



(b) Observed (squares) and fitted (line) variogram for the topographic height.

Figure 3.3: Digital Elevation Model (DEM) for the OPE^3 field site and the variogram used for interpolation of height measurements and generation of the DEM. NAD83 is the coordinate system of the North American Datum of 1983.

This type of subsurface flow forms cascading pathways from one local depression in the clay layer to another, which, although spatially constant, is temporally dynamic. In contrast to the deeper layers, the top soil layer has a larger water holding capacity caused by a higher organic matter content (remains of previous crops) and a higher silt and clay fraction than the deeper layers.

At the Powder Mill SCAN site just outside the OPE^3 field the soil pedon is characterized by the USDA NRCS as coarse loamy, siliceous, mesic Typic Paleudult. Results of their texture analysis at a single point profile are shown in table 3.1. The bulk density increases with depth from 1.61 to 1.75 g/cm^3.

3.3.3 Land use and land cover

Topographic maps, photos and other geographic information system (GIS) datasets about land use in the research area reveal a minor development and change in land use during the past years. However, the field site itself has changed from an area with limited infrastructure, e.g. roads and little buildings, to an agricultural area (figure 3.6 and 3.7, adapted from http://terraserver.homeadvisor.msn.com/, 2003, 2005). During the past years and up till now, each summer, the field has been covered with corn. The 4 sub-watersheds are managed by the BARC-ARS of the USDA with different crop production systems:

- Field A: uniform application of fertilizer in the form of processed liquid manure;
- Field B: conventional farming approach where the field receives a uniform application of agricultural chemicals;
- Field C: from 1998 through 2000: same treatment as watershed A; from 2001 to present: same farming practice as watershed B. Field C has similar soil properties, but the subsurface hydrology is reported to be significantly different from the other three watersheds;
- Field D: precision farming approach where within-field locations are given agricultural chemical applications according to site-specific conditions at the locations.

Table 3.1: Characterization of the soil at the SCAN site and upper 9 cm spatial average over the OPE^3 field.

Depth [cm]	*Horizon*	*Clay* [%] (<0.002 mm)	*Silt* [%] (0.002-0.05 mm)	*Sand* [%] (0.05-2 mm)
0-9	OPE^3	15.62	22.19	62.17
0-14	Ap1	5.7	23.9	70.4
14-29	Ap2	6.3	25.9	67.8
29-46	Eb	6.2	30.8	63.0
46-83	Bt1	8.3	15.7	76.0
83-129	Bt2	7.2	6.2	86.6

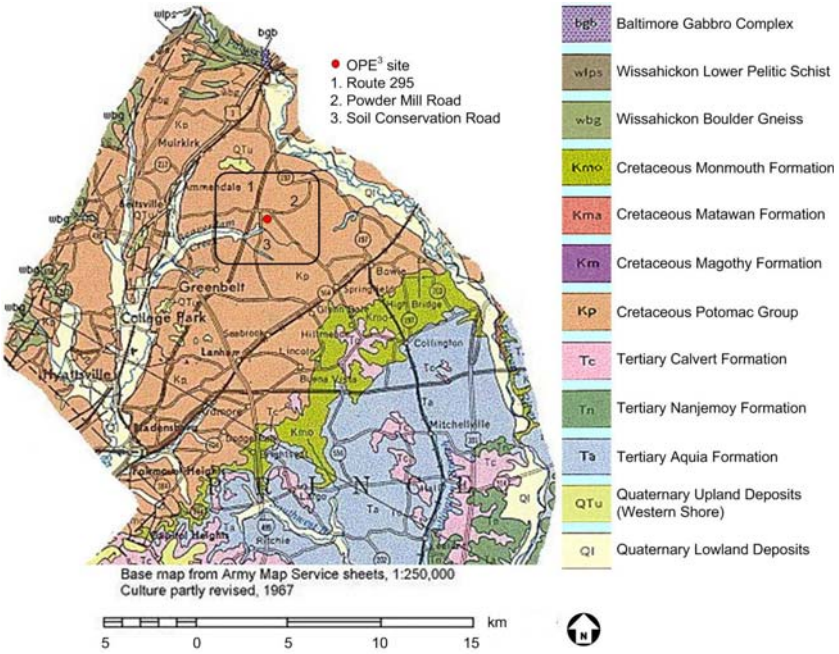
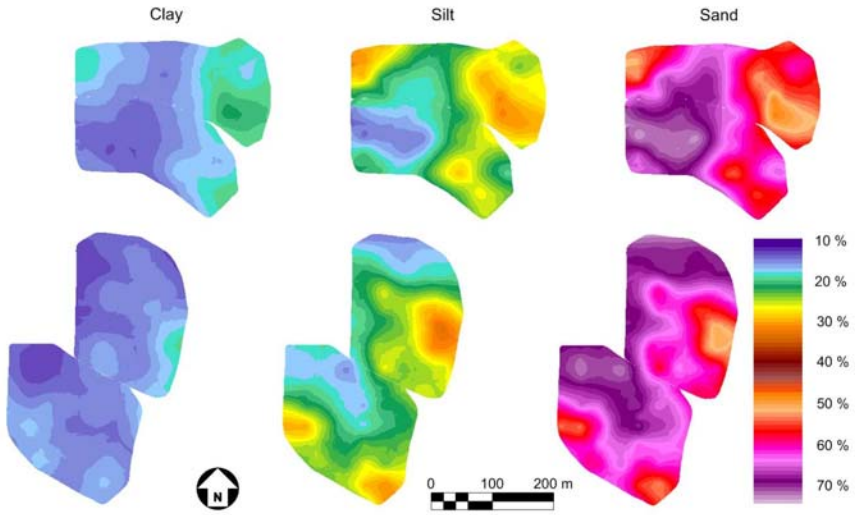
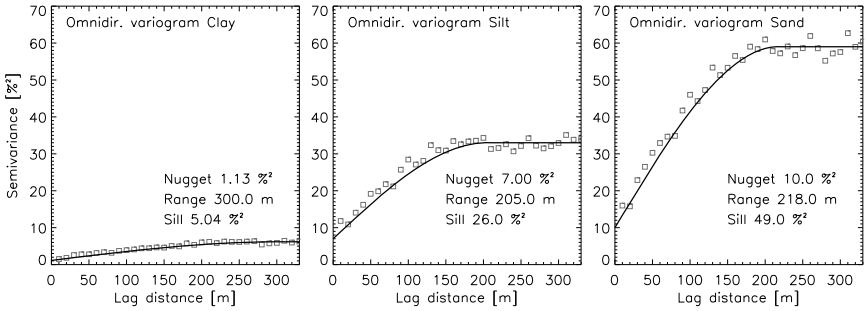


Figure 3.4: Geologic formations in northern Prince Georges County, adapted after Maryland Geological Survey's Geologic Map of Maryland (1968).



(a) Maps of soil texture.



(b) Observed (squares) and fitted (line) variogram for the topographic height.

Figure 3.5: Texture of the upper 9 cm in the OPE^3 field site, observed variograms and corresponding fitted spherical variograms.

During the intensive field campaign in 2002 led by the HSB of NASA/GSFC, corn height was measured almost daily in field A for 21 corn plants. The spatial averaged value and standard deviation are shown in figure 3.8. The monthly mean and the monthly averaged spatial standard deviation are also plotted. The monthly mean heights in 2002 are taken as a first guess for the corn heights in the year 2001 for the model runs that will be described in chapter 5 and further. Additionally, some point measurements taken specifically for study of corn phenology are shown. Physiological corn maturity was noted the first days of October 2001, and the harvest happened on November 1, 2001 without winter crop sowing afterwards. In 2002, the corn maturity occurred earlier, the harvest was on October 1, 2002 and winter cover crop was sown immediately after the harvest.

Spatial information of yield was collected during harvest and was made available for the years 1998, 1999 and 2000. Some statistics are given in table 3.2.

3.4 Meteorological data

3.4.1 Field B

In field B of the OPE^3 corn field, meteorological data are collected. Data from the period June 9, 2001 through September 30, 2002 were made available by the BARC-ARS of the USDA Hydrology and Remote Sensing Lab. These data are collected at a 10 minute interval by instruments on the 10 m high USDA meteorological tower and cover the most important variables to force hydrological models. Information on the sensors is summarized in table 3.3. Most sensors are from Campbell Scientific, Inc. (CSI). In the further course of this text, these meteorological data will be referred to as 'B_meteo data'. These data were used as the basic dataset. Gaps in the time series were filled with the datasets that are discussed next.

3.4.2 Powder Mill SCAN site

There is a meteorological SCAN tower situated just outside (at the border of) field D, at an elevation of 32 m, latitude 39.017 N, longitude 76.850 W (figure 3.9). The height of the tower is about 3.5 m. The vegetation cover at the site is grass. Data

Table 3.2: Yield statistics in liters/ha during 1998, 1999 and 2000 for the OPE^3 field, with minimum min, maximum max, mean m, and standard deviation $stdv$.

Year	*min*	*max*	*m*	*stdv*
1998	142.24	12137.73	4693.99	2152.35
1999	0.00	13770.52	1736.85	1806.47
2000	0.00	27387.57	9724.11	3615.19

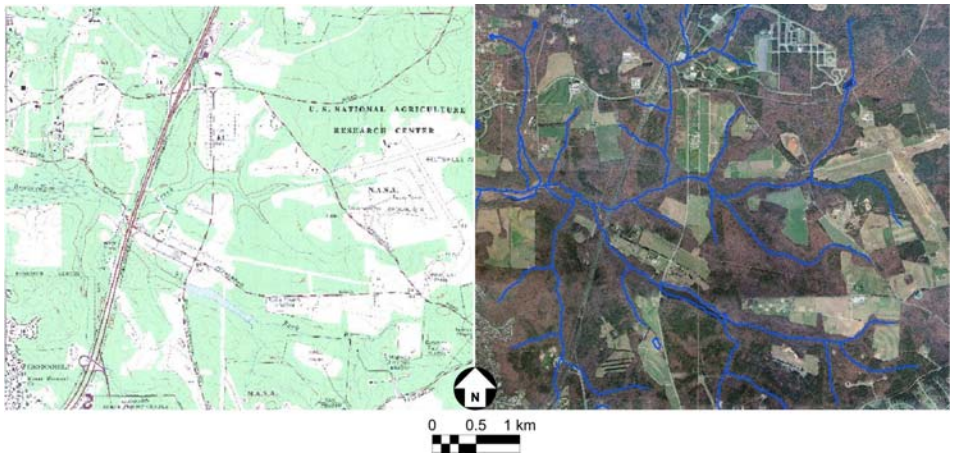


Figure 3.6: Left: topographic map (USGS 1965) of the area around the OPE^3 site. Right: Digital Orthophoto Quadrangle (DOQ) composed from information in the late nineties with a blue vector layer representing rivers in the area around the OPE^3 site (image courtesy of the USGS). The lower left corner is positioned at 76.875 W, 39.000 N.

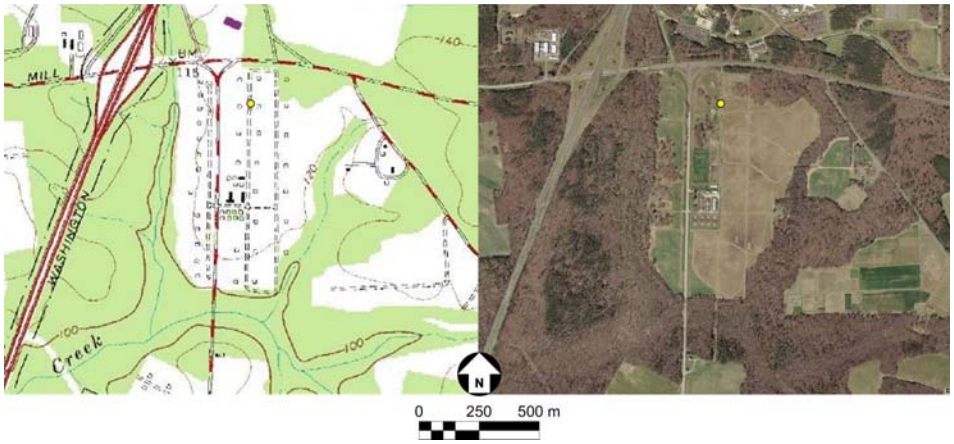


Figure 3.7: Left: detail of a topographic map (USGS 1979) of the OPE^3 site showing some infrastructure at the site. Right: picture of the site at April 7, 2002 (image courtesy of the USGS). The dot in the map is at 76.846 W, 39.033 N.

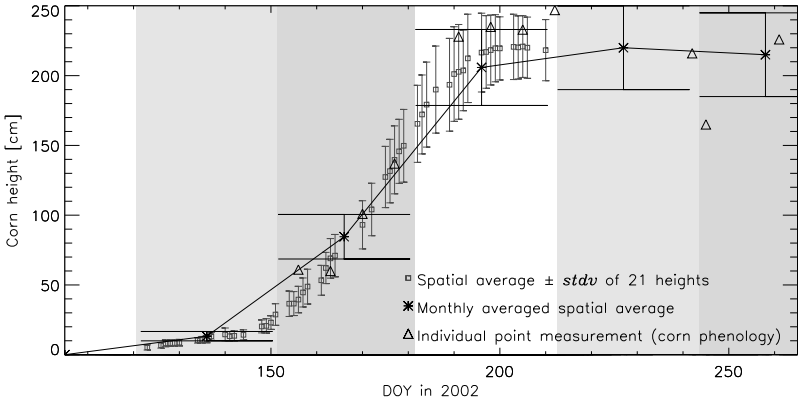


Figure 3.8: Spatial mean corn height and 1 standard deviation (*stdv*) during the summer 2002 in field A of the OPE^3 site. The background shading gives an indication of the months. DOY stands for day of the year.

from this tower are available through the WWW from October, 2001 to present. Data at the SCAN sites are recorded every hour. Information on the sensors is summarized in table 3.3. In the further course of this text, these meteorological data will be referred to as 'SCAN_meteo data'.

3.4.3 Station #3 Old Beltsville Airport

As the SCAN_meteo data do not cover the whole period for which B_meteo data are available, a third closest site, the Station #3 Old Beltsville Airport, was selected to fill gaps in the B_meteo data during 2001. Data at a 15 minute interval from January to December, 2001 were provided by the BARC-ARS USDA Farm Operations Branch. The site has a 3.05 m high tower, situated at a latitude of 39.026 N and a longitude of 76.825 W (figure 3.9). Information on the sensors is summarized in table 3.3. In the further course of this text, these meteorological data will be referred to as 'OBA_meteo data'.

3.5 Processing and analysis of meteorological data

Preprocessing the meteorological data is a first necessary step before use of the data for modeling purposes. Quality control of the data showed missing and/or bad values in all 3 datasets. The raw data were converted to data required for the Community Land Model (CLM), which will be introduced in chapter 5: temperature [K], wind [m/s], specific humidity [kg/kg], incident solar radiation [W/m^2] and total precipitation [mm/s] (units are as required by default for CLM2.0). Other

Table 3.3: Overview of sensors used to measure meteorological data in field B, at the SCAN site and at the Station #3 Old Beltsville Airport. The units of the raw data are given.

Variable (B_meteo)	*Sensor*
Air temperature at 9.5 m [℃]	HMP45C (CSI)
Relative humidity at 9.5 m [%]	HMP45C (CSI)
Wind speed at 10 m [m/h]	Anemometer and vane, R.M. Young's Wind Sentry, 03001 (CSI)
Solar radiation at 5 m [W/m^2]	Silicon pyranometer, LI200X (CSI)
Precipitation at 2.5 m [mm]	Tipping bucket rain gage, TE525WS (CSI)

Variable (SCAN_meteo)	*Sensor*
Air temperature at 2.3 m [℃]	HMP45C (CSI)
Relative humidity at 2.3 m [%]	HMP45C (CSI)
Wind speed at 3.3 m [miles/h]	Propeller type, R.M. Young wind monitor, 05103 (CSI)
Solar radiation at 3.1 m [W/m^2]	Silicon pyranometer, LI200X (CSI)
Precipitation at 2.3 m [inches]	Tipping bucket rain gage, TE525WS (CSI)
Barometric pressure [inches Hg]	Silicon capacitive pressure sensor, CS105 (CSI)

Variable (OBA_meteo)	*Sensor*
Air temperature at 1.5 m [℃]	HMP45C (CSI)
Relative humidity at 1.5 m [%]	HMP45C (CSI)
Wind speed at 3 m [m/s]	R.M. Young wind monitor, 5103 (CSI)
Solar radiation at 3 m [W/m^2]	Silicon Pyranometer, LI200X (CSI)
Precipitation at 1 m [mm]	Tipping bucket rain gage, TE525 (CSI)

forcings are calculated within the CLM code. This mainly included conversion of recorded relative humidity, RH, to specific humidity, SH, and conversion of units. Relative humidity is given by

$$RH = \frac{e_a}{e_s}[-] \tag{3.3}$$

with e_a the actual vapor pressure [Pa]. e_s is the saturated vapor pressure [Pa] and can be calculated by [Tetens, 1930]:

$$e_s = 0.6108\ \exp\left(\frac{17.27\,T}{237.3+T}\right)\cdot 10^3 \quad [\text{Pa}] \tag{3.4}$$

with T the temperature expressed in ℃. The specific humidity SH is given by

$$SH = \frac{0.62197\,e_a}{p - 0.378\,e_a} \quad [-] \tag{3.5}$$

with p the pressure [Pa].

The different meteorological datasets were preprocessed independently. For this research, a 1 hour aggregation level was chosen to compare the data. The datasets

Table 3.4: Overview of the available meteorological data for the OPE^3 field over the period 2001-2002.

Source	*Time resolution*	*Dates*
B_meteo	10 min	June 9, 2001 - September 30, 2002
SCAN_meteo	1 hour	October 30, 2001 - today
OBA_meteo	15 min	January 1, 2001 - December 31, 2001

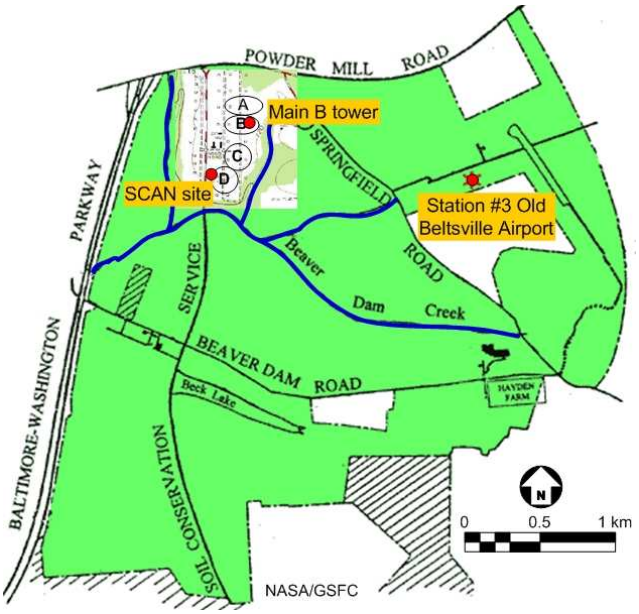


Figure 3.9: Position of the 3 meteorological sites used for the research on the OPE[3] field (image courtesy of USDA-BARC ARS and Farm Operations).

showed a very good similarity in the temporal pattern of the observations. The differences between the SCAN_meteo or OBA_meteo and B_meteo datasets during the overlapping periods were minor. When comparing the data sets, one should be aware that the measurements from the different towers are taken at different heights (see table 3.3)

- The *temperature* is supposed to decrease with height almost linearly in the lowest 10 km of the atmosphere [Garratt, 1992]. It was found that the temperature values for the B_meteo data were only lower for relatively high temperatures, while for low temperatures B_meteo data were slightly higher. The temperature in the SCAN_meteo and OBA_meteo data showed slightly higher peaks and slightly lower minima than the temperature B_meteo data. However, the differences were almost negligible. This resulted in a slightly smaller temporal variation in temperature for the B_meteo data compared to those from the other towers.

- In the lowest parts of the atmosphere, the troposphere, it is well known that the *wind speed* shows a logarithmic increase with height [Garratt, 1992]. This principle was found in the datasets: the SCAN_meteo and OBA_meteo data showed consistently lower values than the observations at 10 m height for the B_meteo data. There was also a higher variation in wind speed for the B_meteo data.

- The *relative humidity* for the SCAN_meteo data was consistently higher than for the B_meteo data and the OBA_meteo data showed also higher values during the months June and July 2001, while for the rest of 2001 the OBA_meteo data showed lower maxima. Since the HMP45C sensors for the SCAN_meteo and OBA_meteo data are situated just above the maximal corn heights and relatively close to the ground, respectively, the higher RH for these sites can be explained by evapotranspiration.

- To convert from relative humidity to specific humidity, *pressure* data were needed (Eq. (3.5)). These are only available for the SCAN site. During the processing of the OBA_meteo data in 2001, the pressure values of the SCAN_meteo data in 2002 were used, while for the B_meteo data for the months June to October in 2001, the monthly averaged pressure from the SCAN_meteo data in 2002 was used as a best guess.

- Around noon, the highest values for the *incoming solar radiation* for the SCAN_meteo data were considerably lower than those for the B_meteo data, except when the radiation was quite low. However, the difference in height of the sensors for this variable is only 2 m and one would not expect a noticeable difference because of the minimal absorption by the atmosphere over this small distance. The instruments are of an identical type, both measuring incoming solar radiation. The OBA_meteo solar radiation observations were almost identical to the B_meteo data. Whenever there was precipitation (overcast weather), the incoming solar radiation decreased significantly.

- For the *precipitation*, the difference in height of the sensors is small. The timing of the precipitation events and their intensities were very similar. In general the intensity of precipitation events from the OBA_meteo data resembled better than the SCAN_meteo data to the B_meteo data. This is probably mainly due to the similar aggregation level of the raw datasets. Of all the meteorological data, the observations for precipitation showed most discrepancies between the datasets, probably due to a combination of the spatial variability of precipitation and the original differences in aggregation times (in recording) of the raw data.

The above discussion highlights the problem of small errors in meteorological forcing data: no measurement is perfect, there are all kinds of errors, due to e.g. different instrumentation accuracy, calibration, influence of terrain and position. Furthermore, while all the discussed variables are spatially variable, point measurements were taken to represent the forcings over a whole field.

Aggregation to hourly time steps allowed to combine information of the three datasets, to complete and to interpolate them where needed. Basically, the B_meteo data at about 10 m were chosen as representative, as the large tower is situated in the corn field B, while the SCAN tower is situated just outside field D in grass and close to a road. Where B_meteo data were missing, SCAN_meteo data were used after converting them using the relationship found by regression of the data

from both sites, because the SCAN site is closest to the B field (see figure 3.9). During the months June through October 2001, when there are no SCAN_meteo data, the gaps in the B_meteo data were filled by converting OBA_meteo data. Regression coefficients and determination coefficients between the SCAN_meteo data and B_meteo data for the overlapping period of October 30, 2001 through September 30, 2002 and those between the OBA_meteo data and B_meteo data for the overlapping period of June 9, 2001 through December 31, 2001 are shown in table 3.5. Corresponding scatter plots are given in figure 3.10.

In general, a high correlation was found between the time series of the B_meteo data and the SCAN_meteo data. For the precipitation there was a slight difference in measuring time: the SCAN_meteo data consist of the current precipitation every hour, while the hourly rate precipitation for the B_meteo data was calculated by averaging 10 minute interval data. Additionally, precipitation is spatially very variable, even at very small scales [De Lannoy *et al.*, 2005], which explains some small time shifts between precipitation events. All this resulted in a very low determination coefficient between precipitation from the B_meteo data and that from the SCAN_- and OBA_meteo data. Therefore, the B_meteo data for precipitation were completed where needed by copying (instead of converting) the SCAN_- or OBA_meteo data. The resulting meteorological dataset used for the modeling is shown in figure 3.11.

3.6 Surface runoff and subsurface drainage

Runoff data for field A and C were reported by Chinkuyu *et al.* [2004] and Chinkuyu *et al.* [2005] and are copied into table 3.6. Field B is similar to field A in many aspects, but field C has seepage zones which are different than in the other 3 fields. Subsurface drainage is difficult to quantify but is probably a major component of the water balance. An approximate estimate (Timothy Gish, personal communication) is that at least 30 % the precipitation leaves the root zone as drainage.

Table 3.5: Regression coefficients α and β and determination coefficients (r^2) for the regression $B\_meteo\ data = \alpha + \beta.(SCAN\ or\ OBA)\_meteo\ data$ between the SCAN or OBA_meteo data and B_meteo for the overlapping periods.

Variable	SCAN α	SCAN β	SCAN r^2	OBA α	OBA β	OBA r^2
Air temperature [°C]	0.5084	0.9681	0.9855	1.0090	0.9494	0.9788
Relative humidity [%]	-2.2496	0.9599	0.9567	0.5747	0.9624	0.9006
Wind speed [m/s]	0.1747	1.3475	0.7994	0.1445	1.1808	0.7947
Solar radiation [W/m^2]	8.1389	1.1897	0.8892	12.5502	0.9281	0.8640
Precipitation [mm/s]	-1e-05	0.4099	0.1881	2e-05	0.1589	0.0236

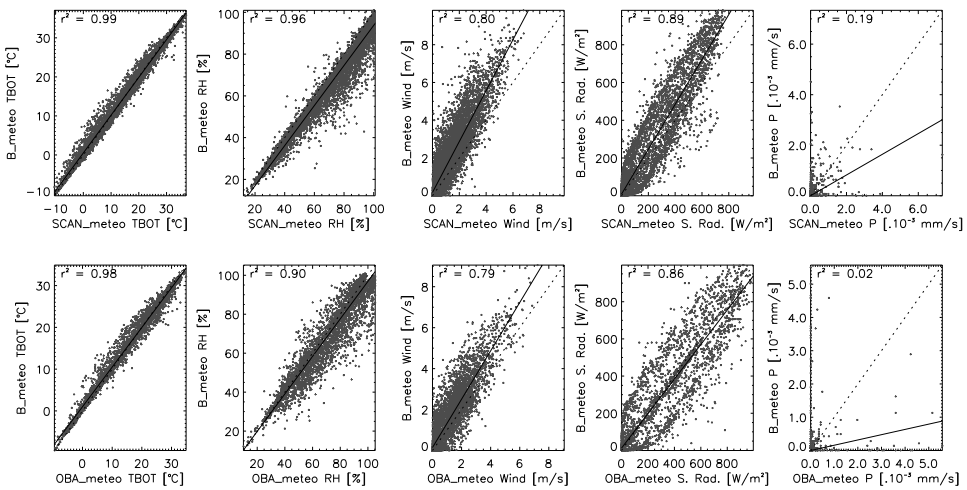


Figure 3.10: Regression between B_meteo data and SCAN_meteo (upper plots) or OBA_meteo (lower plots) data, with TBOT = temperature, RH = relative humidity, S. Rad. = solar radiation and P = precipitation.

Table 3.6: Total yearly runoff [cm] in fields A and C of the OPE^3 field, following Chinkuyu *et al.*, 2004* and Chinkuyu *et al.*, 2005.

Year	1999	2000	2001	2002	2003
A	1.91	0.66 (0.20)*	2.91 (2.06)*	0.98 (0.20)*	10.59
C	6.23	7.71 (6.59)*	7.94 (8.87)*	4.69	14.67

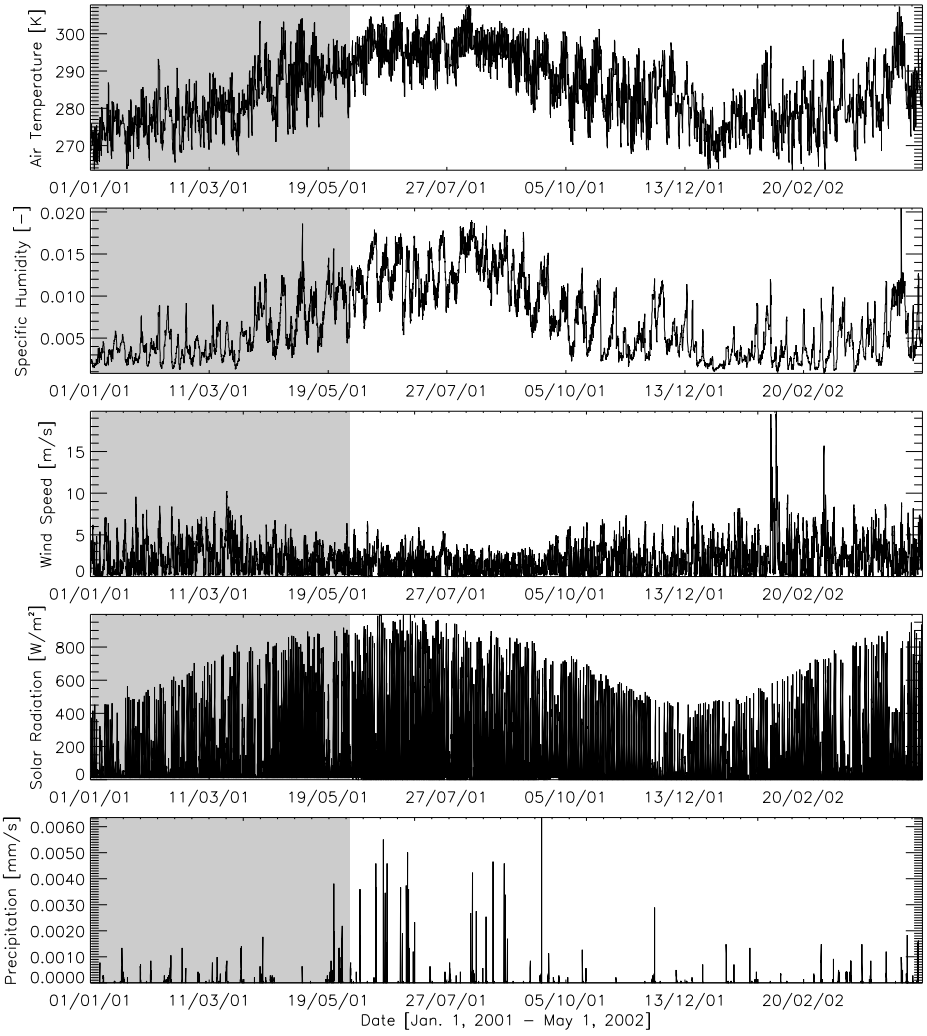


Figure 3.11: Meteorological data used to force the CLM. The data in the gray period are pure OBA_meteo data converted by the regression between B_meteo and OBA_meteo data. B_meteo data are available after this period only.

3.7 Soil moisture data

3.7.1 OPE^3 field A, B, C, D

In each sub-watershed of the OPE^3 field, 12 soil moisture probes have been installed (figure 3.12). Capacitance probes (EnviroSCAN, SENTEK Pty Ltd., South Australia) were installed to independently determine how subsurface restricting layers (clay lens), detected by ground penetrating radar (GPR), influence subsurface soil water dynamics [Gish *et al.*, 2002]. Capacitance probes measure volumetric water contents within a 10 cm radius from the sensor's center [Paltineanu and Starr, 1997; Starr and Paltineanu, 2002]. The site is relatively sandy and receives about 900 mm precipitation each year, which makes that the dielectric constant is affected by neither a high soil surface (fine texture [Roth *et al.*, 1990]) nor by salts. EnviroSCAN is a widely used continuously logging soil moisture monitoring system. Soil moisture values were recorded every 10 minutes.

The probes were named following a 3 digits system. The first letter represents the name of the sub-watershed (A, B, C, D), the second letter (L, H, M) refers to the estimated infiltration at the point of installation (Low, High, Moderate) and the third digit (1, 2, 3, 4) shows the number of the connector cable the probe is connected to in order to transfer data to the logger [Gish *et al.*, 2002]. The general pattern is that 3 probes at points with a different kind of infiltration are connected to 1 connector cable. H-probes have sensors at 10, 30 and 80 cm. L- and M-probes have sensors at 10, 30, 50, 120, 150 and 180 cm. L-probes have an additional sensor at 80 cm depth. Data collected during from May 1, 2001 through April 30, 2002 (and to October 1, 2002 for the sensors in field A, but not processed) were made available by the BARC-ARS of the USDA. During this period probes AL3, AL4, AM3, AM4, AH3, AH4, CL3, CL4, CM3, CM4, CH3 and CH4 were not operational because of technical defects (hit by lightening), causing that 36 out of 48 probes remained operational.

3.7.2 Powder Mill SCAN site

Soil moisture measurements are taken at 1 point at the Powder Mill SCAN site (figure 3.12) at 5 different depths: 5, 10, 20, 50 and 100 cm. The type of sensors used are Stevens-Vitel Hydra probes (Stevens Water Monitoring Systems, Inc.). These sensors simultaneously measure soil moisture, salinity and temperature of the soil. They perform high frequency complex dielectric constant measurements, which resolve the capacitive and conductive properties of the soil and consequently determine the soil moisture and salinity. The sensors need to be installed horizontally and have a precisely defined sensing area of a cylinder of 2.5 cm in diameter and 6 cm in length, bounded by the three outer tines, the probe head and the free ends of the tines. Data from these probes are available over the same period and at the same frequency (every hour) as the meteorological data from the SCAN

site, i.e. from October 30, 2001 to present.

3.8 Analysis of soil moisture data

A decent quality control is a requisite for accurate analyses and therefore, bad data were eliminated, but the dataset for soil moisture was not completed for missing data. A quality control is also needed to assure a Gaussian probability density function (pdf) for the observation errors in the further stage of data assimilation for state estimation [Lorenc, 1995], as will be discussed in chapter 9.

The analysis of soil moisture patterns in time and space aims at the improvement of the understanding of processes in the land surface system and allows for the identification of processes in modeling research. As will be summarized in section 3.8.5, several describing statistics also provide interesting information for state estimation through data assimilation.

3.8.1 Temporal characteristics of soil moisture

Descriptive temporal statistics

Time series of observed soil moisture at all depths for all operative soil probes during the period May 1, 2001 through April 30, 2002 are shown in appendix A in figures 1, 2, 3, 4, 5, 6, 7 and 8 and for selected depths in figure 3.13, 3.14 and 3.15.

To compare data and model results in a further stage a transformation was performed on the observations to change the aggregation time from 10 minutes to 1 hour by temporal averaging. It should be remarked that the summer of 2001 was particularly wet, while data from field A and from the SCAN site reveal that in 2002 the summer was really dry. An extensive time series analysis could be conducted on the data series at hand. However, this study is limited to the analysis of some descriptive statistics for the soil moisture at the individual sensors. Statistics for the different sensors and for different depths are summarized in table 1 in appendix A. Possible relationships between statistics for all sensors in the OPE^3 field are shown in figure 3.16.

Temporal average and standard deviation
The distribution of soil moisture in time was studied. The temporal average and standard deviation give an idea of the overall wetness at a sensor and the response of the soil system to meteorological conditions and influences from neighboring soil volumes. In this section all averages are temporal averages over the 1 year time series for individual sensors. Time averaged information of time series is interesting for calibration and validation of land surface models. Through comparison of temporal statistics from observed to modeled soil moisture time series, model

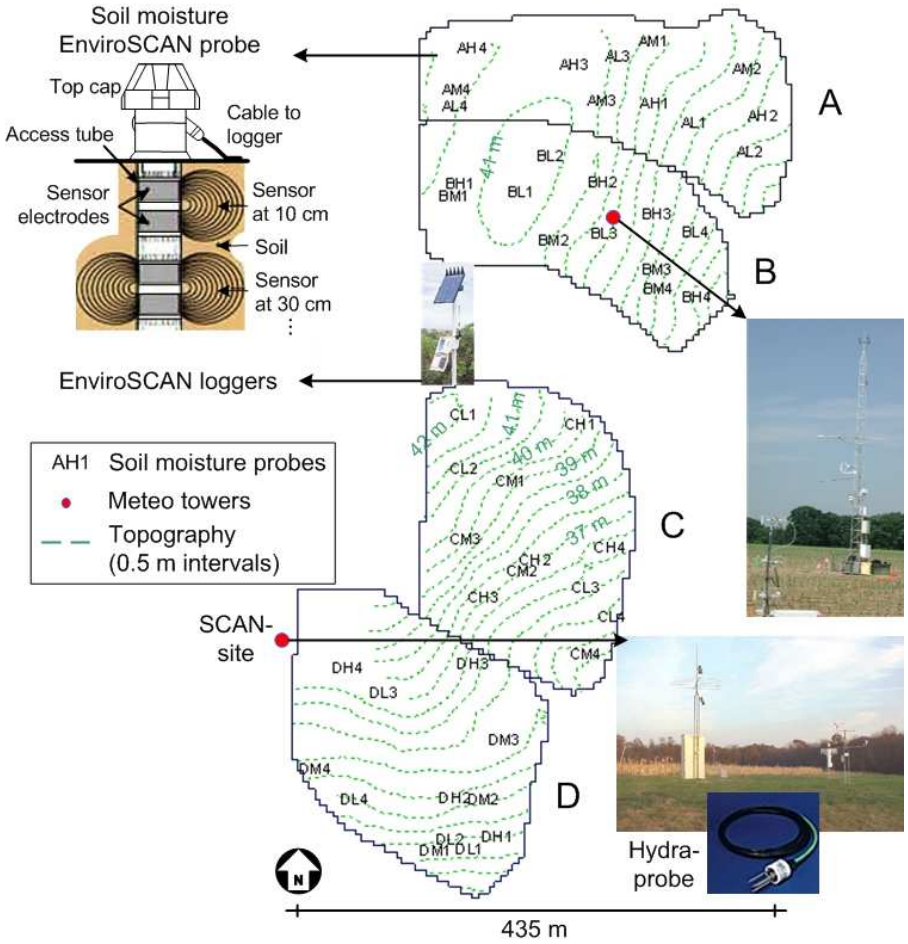


Figure 3.12: The position of soil moisture probes and meteo towers in the OPE^3 field.

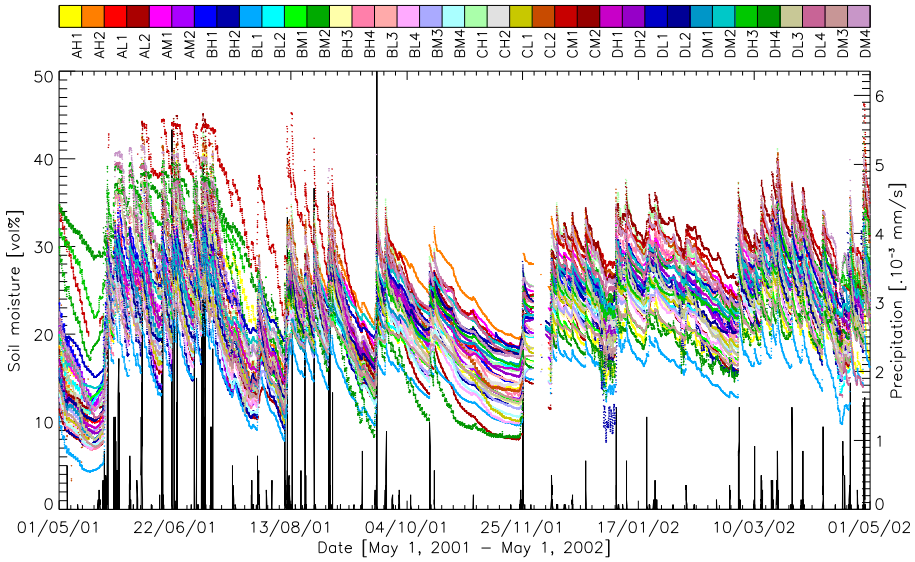


Figure 3.13: Time series of soil moisture for all available sensors at 10 cm depth in the OPE^3 field for 1 year in 2001-2002.

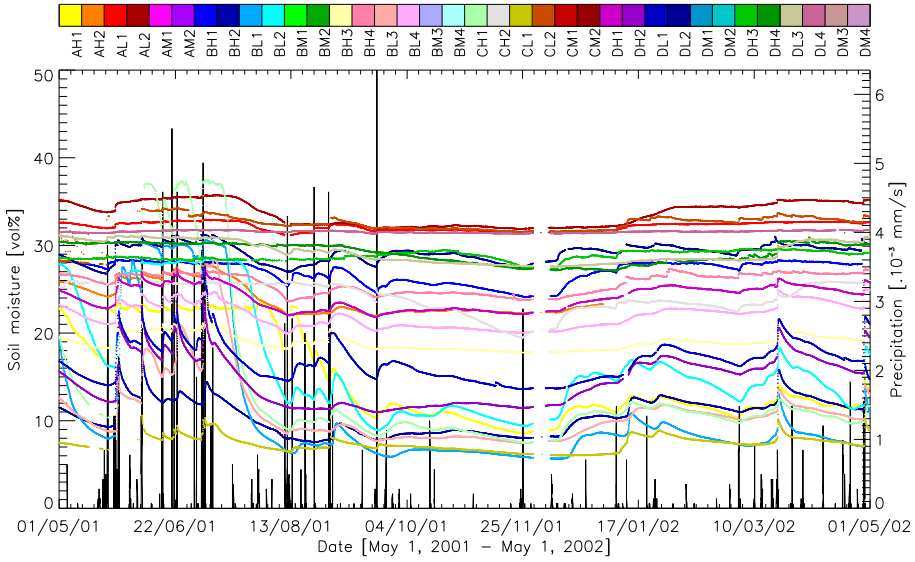


Figure 3.14: Time series of soil moisture for all available H- and L-sensors at 80 cm depth in the OPE^3 field for 1 year in 2001-2002.

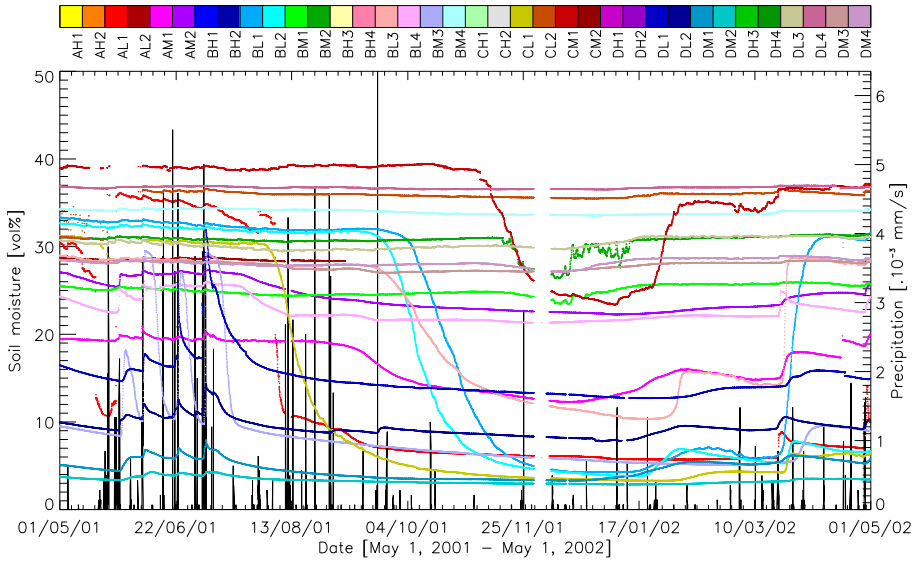


Figure 3.15: Time series of soil moisture for all available M- and L-sensors at 150 cm depth in the OPE^3 field for 1 year in 2001-2002.

parameters can be selected to adjust the model climatology to the observed mean characteristics.

In general, temporal variability of soil moisture is determined by precipitation, evapotranspiration, net lateral flow or vertical drainage, which are referred to as temporal controls. Changes in soil moisture occur at different time scales and are dominated by varying controls. Wilson *et al.* [2004] tried to quantify the relative importance of temporal controls and concluded that the temporal variability of soil moisture in the top 30 cm of the soil profile can be explained by seasonality and individual events and that it is dependent on climate and soil water storage. Furthermore, it was found [Western and Grayson, 2000; Chirico *et al.*, 2003] that time-varying soil parameters, such as dynamic changes in cracks and macro-pores, contributed to variability in soil moisture, particularly during times of rapid change in average soil moisture.

A compacted measure of temporal variability is the standard deviation of soil moisture in time, calculated for individual sensors. From the first column in figure 3.16, it is clear that this statistic decreased with depth. A higher temporal variability is found in the upper layers than for the deeper layers. The upper layer is in direct contact with the atmosphere and there is direct water exchange via precipitation and evapotranspiration and consequently, the standard deviation is large. However, as the top soil layer is exposed to a spatially constant atmospheric forcing and is little influenced by groundwater, the spatial range in the temporal average soil moisture over the test site was smaller than what was observed for

deeper layers.

In deep soil layers (150 and 180 cm) soil moisture nearly remains the same (i.e. wet when the groundwater table resides near the considered layer or dry if soil water drains to deeper layers) resulting in low standard deviations. However, at some sensors, the groundwater table temporarily fluctuates due to the build up of a perched water table caused by preferential (funnel) flow [Gish *et al.*, 2002] followed by a dry-out or slow orderly emptying (i.e. drying out at 120 cm first, then at 150 and finally at 180 cm), probably as a response to evapotranspiration or at some places through leakage through the clay layer. Since at these sites the soil moisture switches a couple of times during a year between two states (nearly saturated water storage and very dry, which is clearly demonstrated in figure 3.15, e.g. for probes BL1 and BL2), the standard deviation of the moisture contents, calculated on an annual basis will be large. Because for several profiles the lower layers easily drain due to their large granular structure, lower time averaged moisture contents are found there, whereas for other places with a relatively constant shallow groundwater table or a less coarse granular soil structure, wetter time averaged conditions may occur due to capillarity. For the measurements at 10 cm depth, the standard deviation is higher for sensors at wet locations than at dry locations, while the opposite is true for the deeper layers.

Autocorrelation

Through calculation of the autocorrelation for increasing time lags of the daily averaged time series of soil moisture for all sensors at all depths, the characteristic time scale was determined. See figure 3.17 for an example of these functions at different depths for probes AM1 and BL4, which are quite representative (see section 3.8.4). It was found that, on average, the autocorrelation reaches an almost constant value (i.e. shows stationarity) for time lags of approximately 50 days (i.e., visually where the graph is flattening or numerically by three times the temporal autocorrelation length, see below) for soil moisture at 10, 30 and 50 cm depth. An increase of the characteristic time scale for deeper layers was observed, with 100 days at 80 cm depth and 130 days or more for 120 and 150 cm depth. Yet, for individual sensors, large deviations towards these values can be found. This can be attributed to the different hydrological processes, e.g. preferential flow, which occur at each of the sensors. The increase in the characteristic time scale for deeper soil layers is very plausible, since the upper layer soil moisture is strongly influenced by the stochastic precipitation input and evapotranspiration, whereas for deeper soil layers, soil moisture changes more slowly. In table 3.7, some statistics of the correlation length are shown for all available sensors at the different depths. The correlation length was determined as the time lag for which the correlation became 1/e (e=2.71828...). The correlation length is about 2 weeks in the rooting zone (10 and 30 cm) and up to 2 months in the deepest layer, although a large variation on these values was observed. This agrees with the characteristic time scale that Skøien *et al.* [2003] found from temporal variograms for soil moisture averaged over the upper 30 cm.

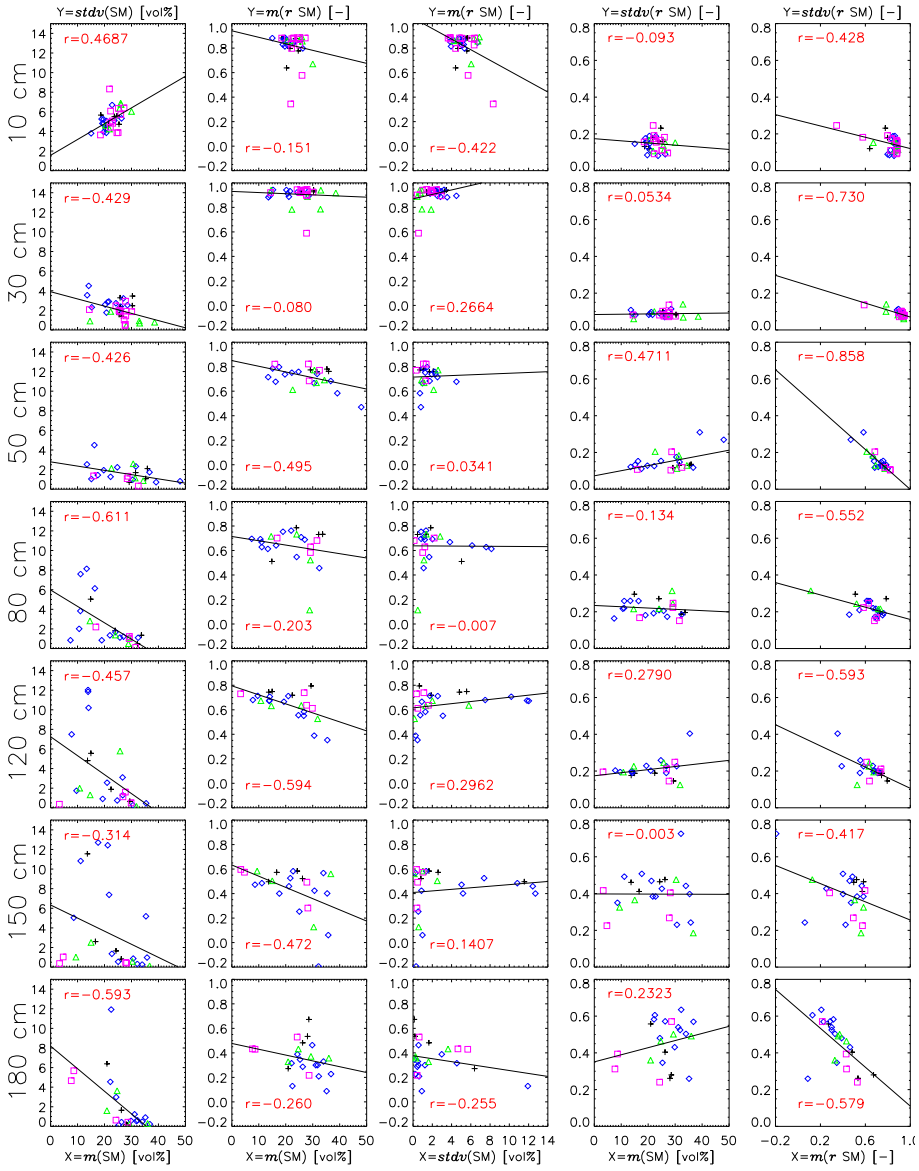


Figure 3.16: Relationship between different statistics for all sensors in the OPE³ field during the period May 1, 2001 through April 30, 2002 and the corresponding best regression with indication of the correlation coefficient r of the regression in red. SM stands for soil moisture, m is for the temporal mean and $stdv$ for the standard deviation. The zero-time lag temporal correlation is given by r and $m(r\,\mathrm{SM})$ and $stdv(r\,\mathrm{SM})$ are the average and the standard deviation of the correlations of a sensor with all other sensors. (Black: field A, blue: field B, green: field C, pink: field D)

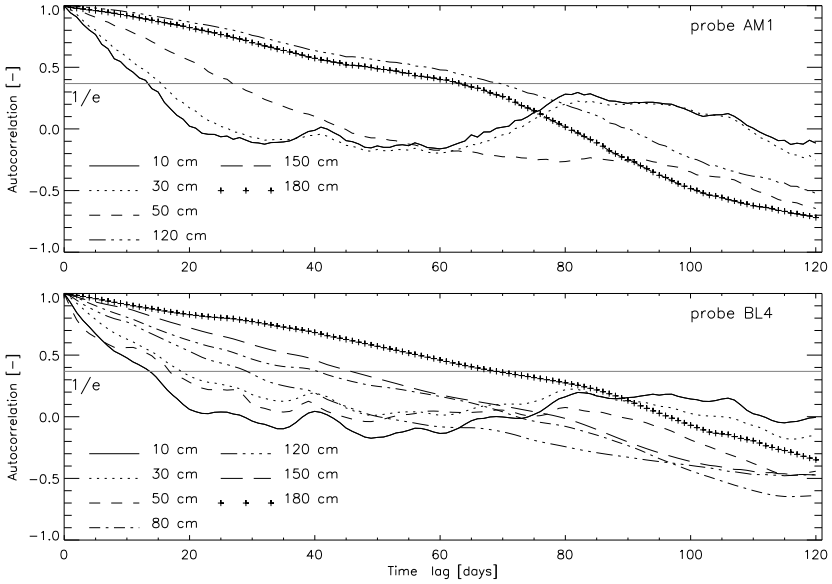


Figure 3.17: Temporal autocorrelation of soil moisture for probes AM1 and BL4 at all depths. The autocorrelation function at 150 and 180 cm depth are overlapping for probe AM1.

Table 3.7: Median M, minimum min, maximum max, mean m and standard deviation $stdv$ [days] of the temporal autocorrelation length, determined for all available (N) sensors per depth.

Depth	10 cm	30 cm	50 cm	80 cm	120 cm	150 cm	180 cm
M	16	16	18	36	45	46	71
min	8	11	14	17	16	20	19
max	59	54	61	77	98	117	82
m	16	19	26	40	51	57	58
$stdv$	10	9	13	15	20	24	21
N	36	36	24	24	24	23	22

Relation to terrain features

For modeling purposes, it is advantageous to link differences in terrain to the variability in soil moisture. Terrain indices are often used to get an idea of the spatial distribution of soil moisture in a simplified way, but are based on the assumption that the controlling process on soil moisture distribution is subsurface lateral flow. The availability of DEMs led to the use of indices like the topindex [Beven and Kirkby, 1979] as common predictors of soil moisture patterns, even in situations where terrain may not be the dominant controlling factor of the soil moisture pattern [Western *et al.*, 1999; Grayson and Western, 2001]. The topindex

or topographic index is given by:

$$\lambda = \ln\left(\frac{A}{\tan\beta}\right) \tag{3.6}$$

with A the area of the upslope watershed per unit contour length, and $\tan\beta$ the local surface topographic slope. The method of Jenson and Domingue [1988] was used to determine the directions of the flow. In figure 3.18 the topindices calculated for different grid cell sizes are shown. Notice that at the borders of the field, the topindex values could not be calculated properly. Wilson *et al.* [2004] stated that it is rare for terrain indices to explain more than half of the spatial variability in soil moisture. One could be in a fortunate modeling situation where terrain indeed does explain most of the variability, but from a static index it cannot be expected that it continuously represents all details of dynamic processes. In earlier research, Western *et al.* [1999] found that some terrain indices performed better in dry conditions, while other indices were advised for wet conditions. A more dynamic index has been proposed by Barling *et al.* [1994], but it still is based on the assumption that subsurface flow is the dominant controlling factor. Furthermore, Wilson *et al.* [2004] found that the spatial distribution of soils and vegetation are of similar importance to that of topography in controlling the spatial distribution of soil moisture.

The relationship between soil moisture for all sensors at all depths in the OPE^3 field and texture, elevation and topographic index λ at all sensor locations was studied. Texture, elevation and topindices were derived from smooth spatial continuous maps of spatially interpolated texture and heights (DEM). When extracting terrain features from 5 m resolution data, one would expect a reduction in variability because of averaging of very little scale variability for texture and height, compared to the 1 m resolution data. There was a very strong relationship between texture or height extracted from a 1 m resolution map and that extracted from a 5 m resolution map (data not shown). However, for the topographic index in figure 3.18, there was barely a relationship between those values obtained for a 5 m resolution DEM and those from a 1 m resolution DEM, which showed a high variability in topindices. Topographic indices at 5, 10 and 20 m resolution for the different sensors were more similar. The high dependence of topindices on scale renders its use as a robust indicator for soil moisture patterns questionable. Furthermore, the soil moisture measurements are representative for a limited (point) volume only, while topindices were calculated over grid sizes of a (few) squared meter(s).

It was not possible to conclude that time averaged statistics (mean, standard deviation, spatially averaged cross-correlation and standard deviation on it) of soil moisture for any probe were in any way dependent on topography (and topographic index) or surface texture in the OPE^3 field.

While there was no significant relation between time averaged soil moisture and terrain features, the relationship between terrain features and the spatial pattern of soil moisture at each time step was often more pronounced. A multiple regression

between soil moisture (y), height (x_1), texture (x_2, x_3 and x_4 for sand, silt and clay, respectively) and topindex (x_5) at 10 m resolution was calculated at each time step:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 \tag{3.7}$$

with a_0 through a_5 the tunable coefficients. This method basically weights the different terrain variables against observed soil moisture for a given occasion. The correlation coefficients for the multiple regressions with and without inclusion of the topindex are plotted in figure 3.19.

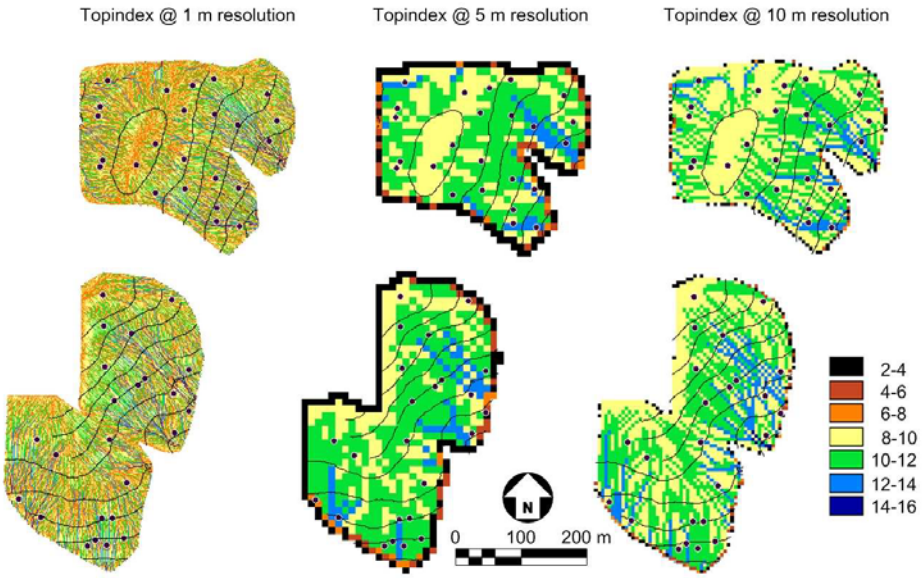
The results highlight that at individual time steps an important part of soil moisture can be explained by a combination of these terrain features, with only a very limited contribution from the topindex, except during the wet summer of 2001. For most depths, the multiple correlation coefficients increased with decreasing soil moisture, which means that it was easier to find a relationship under dry conditions. This is logical, as the soil characteristics and land surface characteristics mainly influence the soil moisture state in dry conditions, while the soil moisture pattern for wet conditions is initially determined by the pattern of precipitation. Multiple correlation coefficients of approximately 0.5 at 10 and 30 cm depth were found, values around 0.8 at 50 cm, and decreasing values at deeper layers, except for 180 cm depth, where multiple correlation coefficients higher than 0.8 were found. The multiple linear correlation coefficient varied in time with soil moisture.

Spatial maps of soil moisture

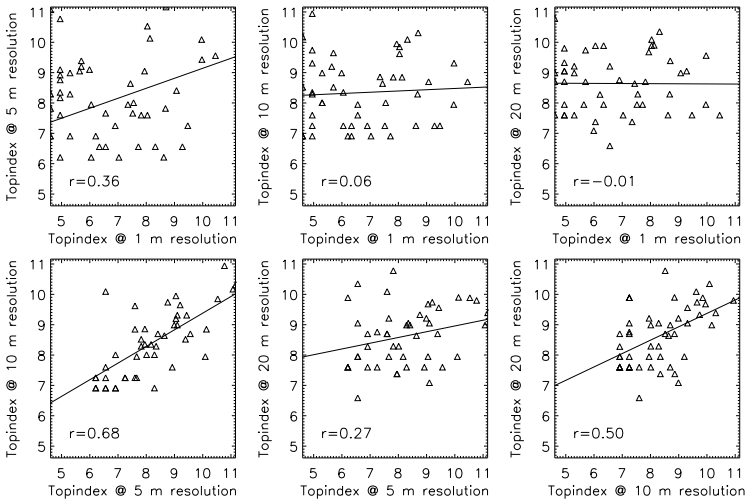
The multiple regression expression found for each time step can be used to generate spatially interpolated maps of soil moisture. This would cause a variability in soil moisture that is completely described by terrain features. Another option is to interpolate the available point measurements by a kriging algorithm. In that case an a priori found correlation structure (see section 3.8.2) must be imposed. Reasonable maps can be generated by combination of both alternatives. A kriging algorithm was used to obtain a smooth interpolated map of soil moisture, using an exponential variogram model (choice based on visual appreciation of the experimental variograms at some time instants), with parameters based on spatial analyses that will be discussed in next sections. A weighted average of the maps was produced, using the uncertainty of the spatial estimates (regression coefficient for the multiple regression and uncertainty of interpolation for kriging) as weight factors. As an illustration, some results are shown in figure 3.20. While a time-variable multiple regression was used, the correlation structure used for kriging was kept time-invariant with a range of 100 m, a nugget of 0 vol%^{2} and a sill of 20 vol%^{2} for all depths. Spatially interpolated plots provide a possible means for visual comparison of observed spatial patterns with modeled ones.

Cross-correlation between time series

To estimate the degree to which two time series of soil moisture are correlated, the cross-correlation with zero time lag was calculated over the complete time series



(a)



(b)

Figure 3.18: Variability of topographic indices for different resolutions. (a) Spatial distribution of the topographic index for the OPE^3 site at different resolutions, with the position of the soil moisture probes indicated as dots. (b) Scatter plots of topographic indices at different resolutions.

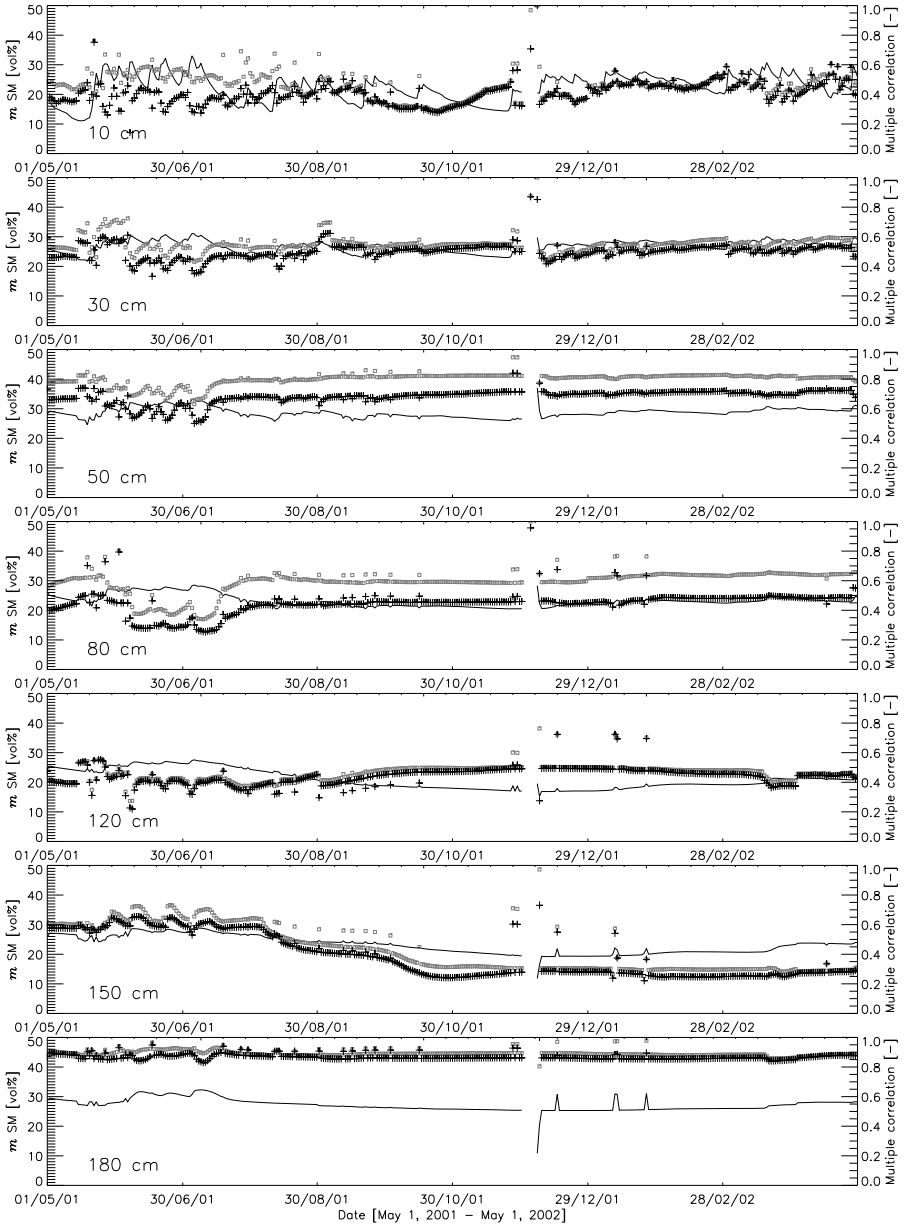


Figure 3.19: Spatially averaged soil moisture, m SM (line), and multiple linear correlation coefficients (gray □) with and (black +) without inclusion of the topindex for the OPE^3 field from May 1, 2001 through April 30, 2002 at different depths. Spikes occur when data from some probes were missing. Gaps occur where no data were available.

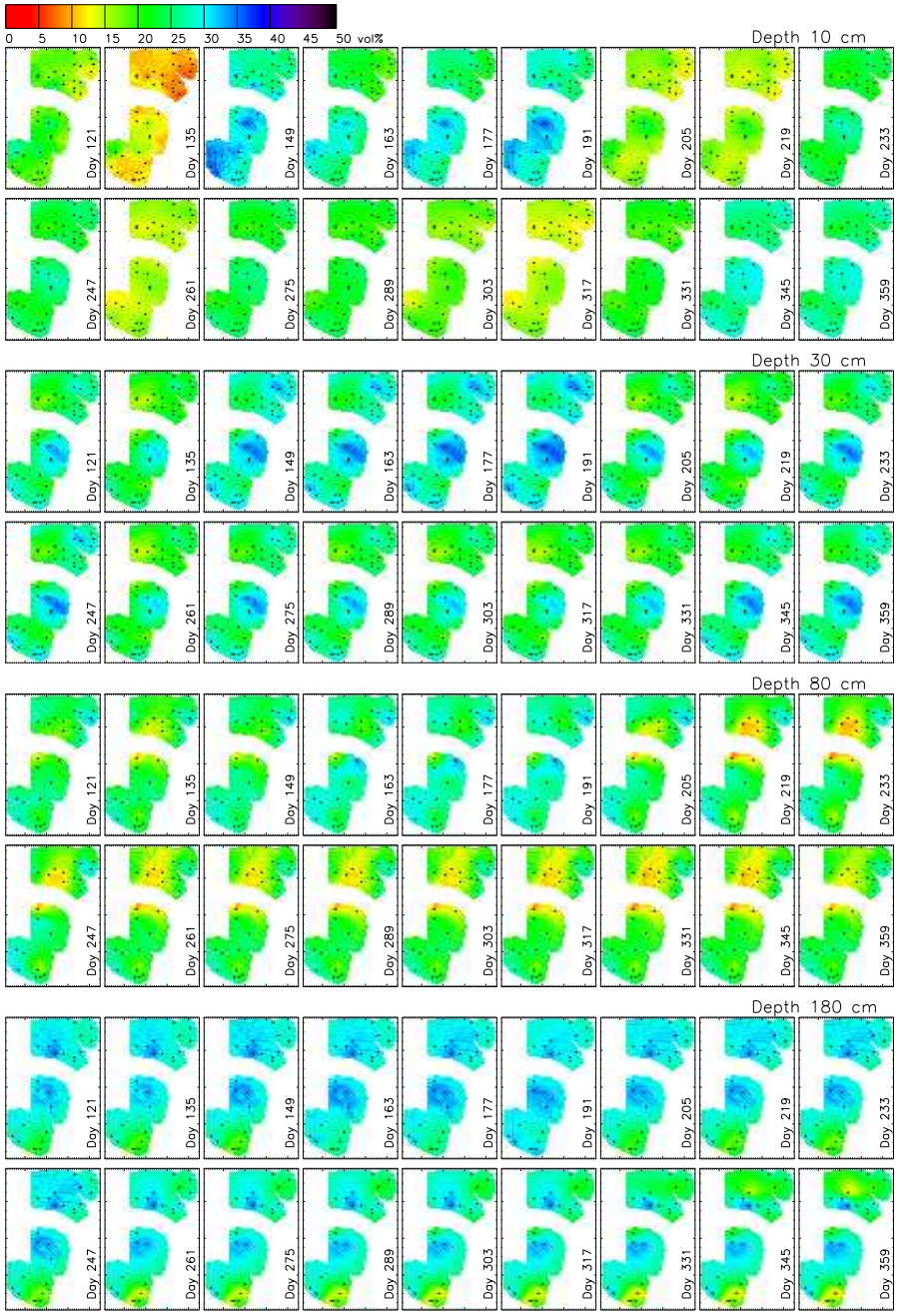


Figure 3.20: Spatially interpolated soil moisture at different layers in the OPE^3 field every 2 weeks, starting at May 1, 2001. Day numbers are counting from January 1, 2001. Black dots indicate where measurements are available.

of 1 year:

$$r_{z_k z_l} = \frac{E[(z_k - < z_k >)(z_l - < z_l >)]}{\sqrt{E[(z_k - < z_k >)^2]\ E[(z_l - < z_l >)^2]}} \tag{3.8}$$

with $< z_k >$ the temporal average of z_k, $E[(z_k - < z_k >)(z_l - < z_l >)]$ the covariance between variables z_k and z_l and $E[(z_k - < z_k >)^2]$ the variance of z_k. z_k and z_l represent (time series of) soil moisture values from 2 different sensors. Only cross-correlation between sensors at the same depths is discussed. Information on the cross-correlation between soil moisture at different points in space is of particular interest for the model structure. A model can explicitly account for cross-correlation, e.g. by allowing horizontal redistribution of water within the transition function. If the model does not include this correlation, the model error will contain cross-correlation. For state updating, the knowledge of cross-correlation between points is very interesting to understand the propagation of information from one point to another through time after state updating or for data selection at the analysis time step. The computational load of state updating can be limited by considering only those points that are highly correlated with the observation point, which is by default the case for assimilation schemes which only use observations within a predefined radius (and hence high cross-correlation), like optimal interpolation, successive corrections or nudging. A reduction of the computational load for methods like Kalman filtering can be introduced through localization of the a priori state error covariance $\mathbf{P}_i^-$ (see later).

Figure 3.21 shows the correlation matrices at different observation depths for the entire year time series. Figures 3.22 and 3.23 show the correlation matrices at some observation depths for the first half of the time series ('summer': May 1, 2001 - October 30, 2001) and the second half ('winter': October 31, 2001 - April 30, 2002), respectively, to discern periods with and without vegetation. In general the correlation coefficients are quite high. It happened that the cross-correlation calculated over the entire year of data was much less than the value calculated for the summer or the winter individually: this is caused by the fact that the total series was made up of different events, with different relationships between the sensors for different time periods. During the winter, probes CM2 and AH2 did not have a representative amount of pairs, while all other probes had about the same number of available data points. During the summer, the number of corresponding time steps for each pair of sensors varies in between the different fields, just as for the whole time series (see table 1 in appendix A).

The correlation between sensors decreased with depth. Note that high cross-correlations mean that sensors react similarly on inputs and that no side effects influence the soil moisture response at certain locations. Low correlations indicate that the soil moisture response at some times may be different because of varying conditions, including depth to the clay layer, occurrence of preferential flow, existence of a perched water table, etc., that induce different responses. This is illustrated for the sensors at 10, 80 and 150 cm depth in figures 3.13, 3.14 and 3.15, respectively. For 10 cm depth, all sensors responded similarly, as their moisture content was mainly influenced by atmospheric forcing. At deeper depths, there is

less similarity in particular during the wet summer of 2001, as they were subjected to different hydrogical conditions (e.g., depth to the water table) and processes (e.g., preferential flow).

Grayson *et al.* [1997] studied the similarity in response of different points in space to distinguish between local and nonlocal controls in spatial patterns of the upper soil layer through time. For the OPE^3 field it is not possible to select periods of local and nonlocal control at the surface layers. However, the reasoning of Grayson *et al.* [1997] originally applied to the upper layer to study the changes in time can be extended to study the difference in controls over a soil profile. Probably, soil moisture in the upper layer in the OPE^3 field is mainly controlled by vertical fluxes (local control) and deeper layers are affected by lateral flow (nonlocal control). The findings for the OPE^3 field indicate that soil moisture patterns obtained from remote sensing data for the surface layer may not be representative for the underlying soil moisture pattern, revealing the need for models and for in situ profile measurements to obtain the underlying information [Hoeben and Troch, 2000; Walker *et al.*, 2001b].

Spatially averaged cross-correlation and standard deviation
One row of the correlation matrix can be plotted as a spatial plot with correlation values relative to one point (sensor) in space as shown in figure 3.24. However, since only some points in space are available, spatial interpolation of the cross-correlation was needed for visualization. From such plots the change in correlation between sensors with increasing distance from a fixed point can be examined, but the spatial structure of the correlation coefficients is difficult to evaluate in areas with limited sensor density.

Average correlation is defined here as the average of the correlations of a time series from a sensor with time series of all other sensors and is found to decrease with depth, i.e. the spatial average over the point values used to generate the maps discussed above. The standard deviation in correlation represents the spatial variation in correlation with different sensors. In the 2nd column of figure 3.16, there was no evidence that a sensor would be more correlated with other sensors if the conditions were wetter at that sensor. There was rather a little decrease in correlation with increasing wetness at a sensor most clearly for 120 and 150 cm depth. This again can be attributed to the dry-out phenomena discussed above, which caused high variability in the correlation between sensors. Further, as can be seen in the 3rd column of plots, there was no evidence that a sensor which shows a particular high or low temporal variation (*stdv*) would be on average more or less correlated with other sensors. The 5th column shows that a lower averaged correlation was related to a higher (spatial) variability in correlation with the other sensors.

It can be stated that sensors with a high average correlation and a low standard deviation on it are typically showing a behavior that is very similar to most sensors and may be identified as representative. This is interesting for sampling purposes as will be discussed further in section 3.8.4. In figure 3.25 all sensors were ranked

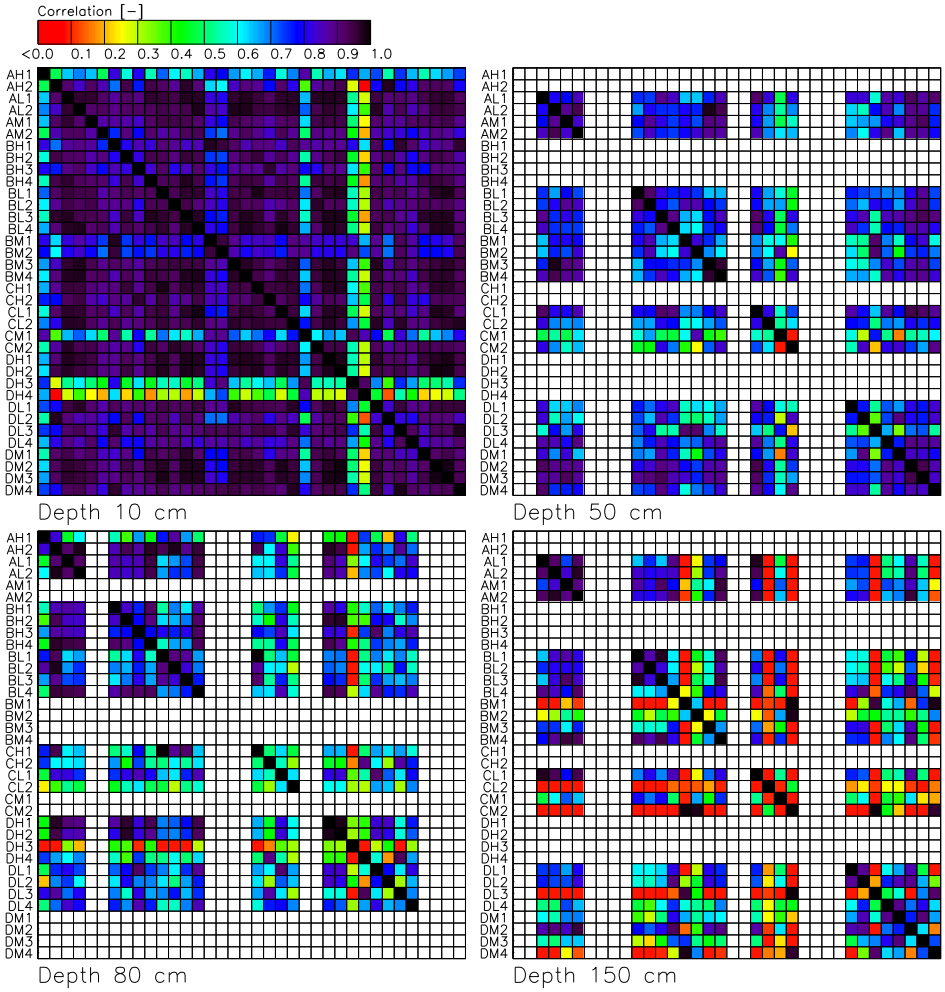


Figure 3.21: Correlation matrix at some observation depths of all sensors for the complete time series of 1 year in 2001-2002. White is for missing sensors.

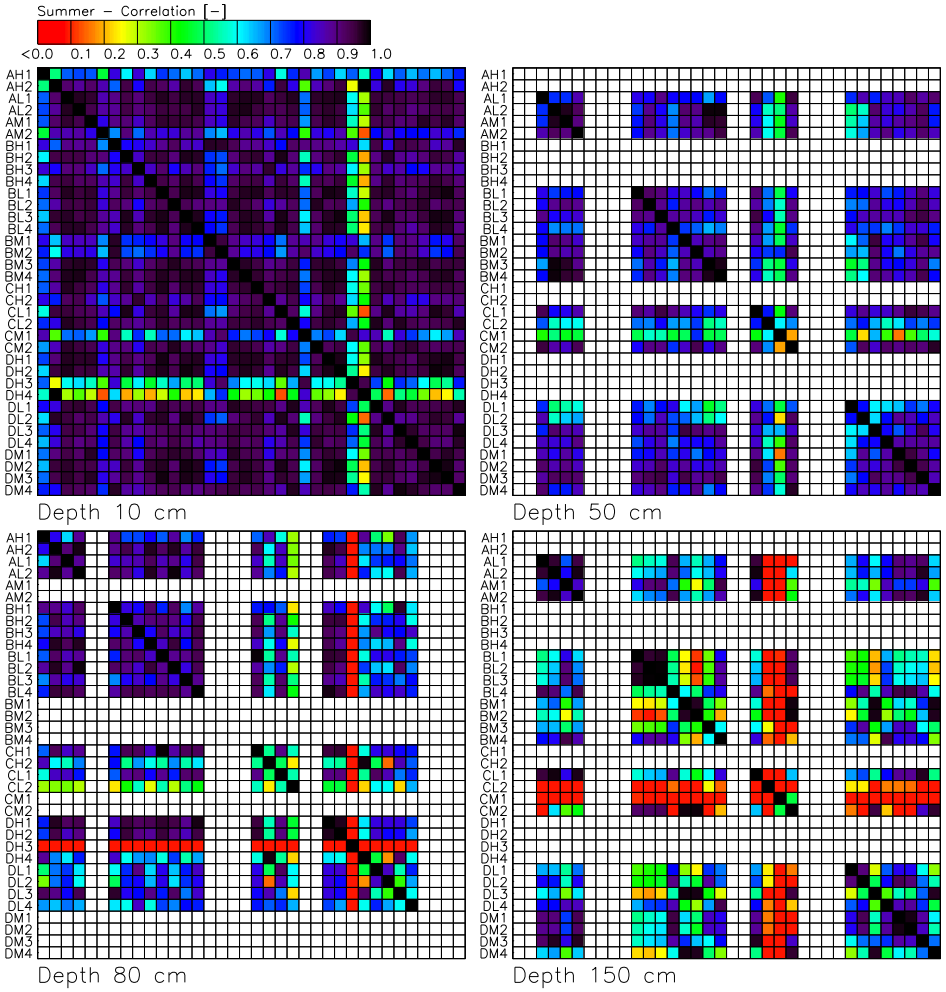


Figure 3.22: Correlation matrix at some observation depths of all sensors for the summer in 2001. White is for missing sensors.

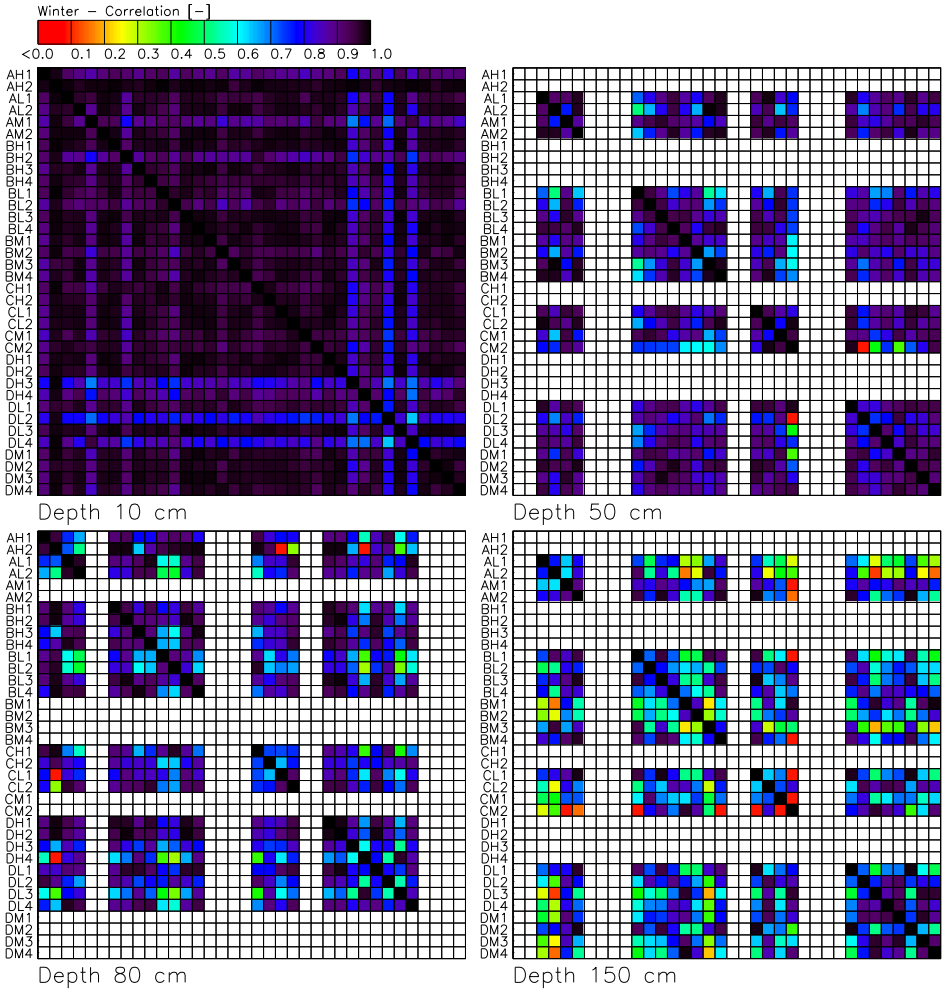


Figure 3.23: Correlation matrix at some observation depths of all sensors for the winter in 2001-2002. White is for missing sensors.

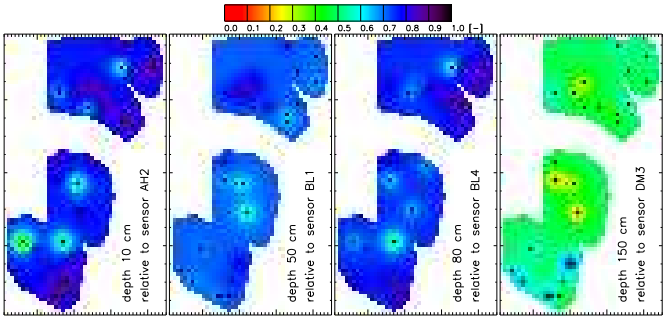


Figure 3.24: Cross-correlation between a given sensor in space and all other sensors.

by their spatially averaged cross-correlation. Also one standard deviation is shown. Sensor DH1 was highly correlated with all other sensors for all its measurement depths (10, 30 and 80 cm), AL2 was highly correlated with most other sensors at deeper layers (80, 120, 150 and 180 cm) and sensor CL2 was highly correlated with other sensors for soil moisture at the surface layer, while it had a very low space averaged correlation at deep layers (80 to 180 cm depth). For the surface soil moisture, sensor DH4 was characterized by a very low space averaged correlation.

Cross-correlation vs. distance and differences in terrain features
Figure 3.26, 3.27, 3.28 and 3.29 show for different periods the relation between the cross-correlation between time series of sensors, with zero time lag (cfr. supra) and the distance, height difference, texture difference, moisture difference between sensors and difference in topographic indices at several resolutions. As stated already, knowledge of the cross-correlation is of major importance in data assimilation applications. If the cross-correlation between points can be linked to a change in terrain features, this is even more interesting. Until now, mostly the relation between cross-correlation and distance was explored for data assimilation schemes, in the sense that only observations within a given radius (which implies with a high enough cross-correlation) have been selected for state estimation at a given point.

As found earlier for averaged correlations, it is clear from the figures that the correlation between sensors decreased with depth and that the variability on the correlation coefficients increased. Overall, there was a decrease in cross-correlation for increasing terrain differences between sensors, even though it was not always very clear due to the large scatter and the relatively limited amount of points. For the individual sub-watersheds A and C (data for individual sub-watersheds not shown), it was less obvious to find a visual relationship between correlation and distance or terrain features: this was caused by the lower amount of observation points in these sub-watersheds. Fields B and D clearly showed a decrease in correlation with increasing distance between sensors. The decrease in correlation with increasing difference in height was very clear in field D and in field B this

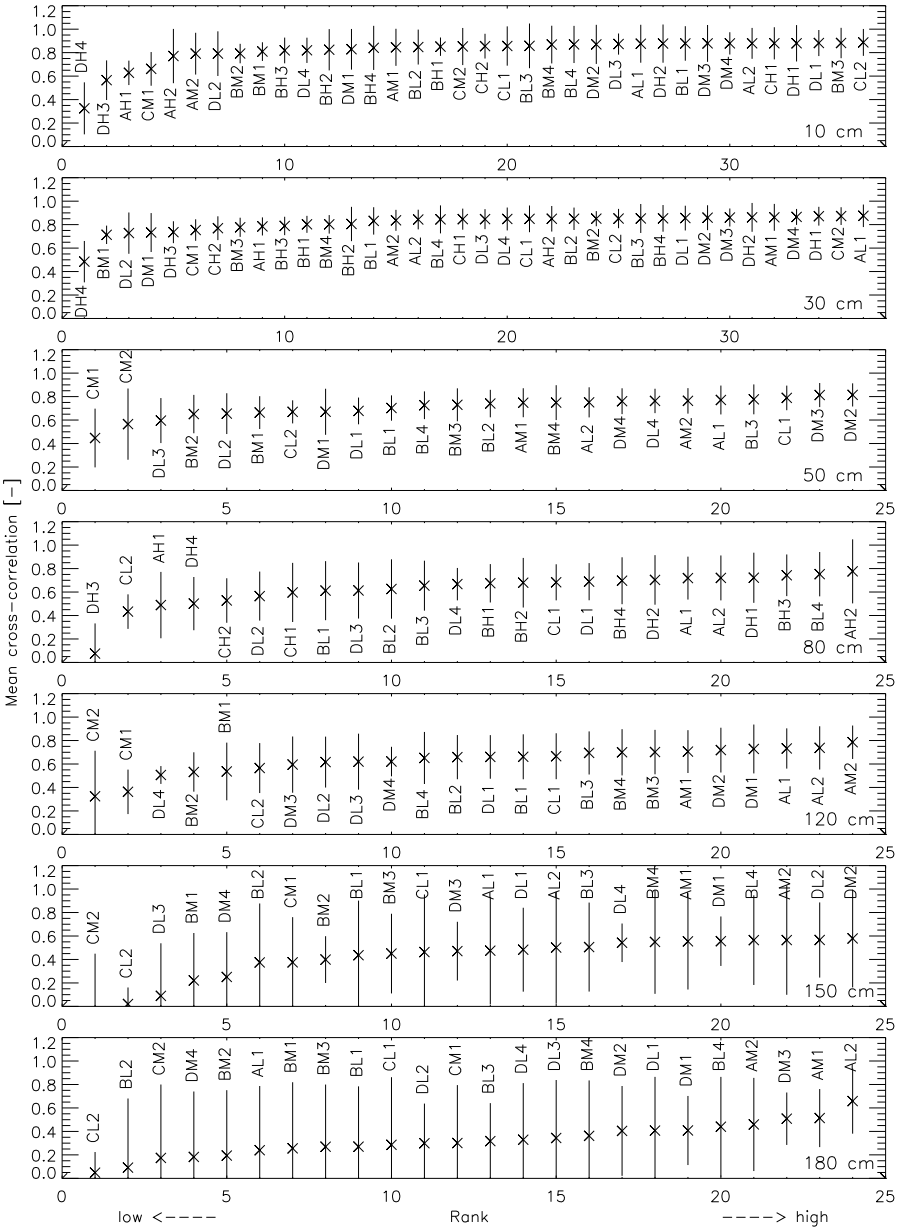


Figure 3.25: Ranked averaged cross-correlation and standard deviation for all sensors in the OPE^3 field.

was most clear at 50 cm depth. The absolute values of the correlation coefficients of the regression lines were slightly smaller than for the relation with the distance. For the complete OPE^3 field the correlation coefficients for the regressions were lower than for the individual watersheds, but the same trends could be found.

Figure 3.29 shows no reasonable decrease in correlation with increasing difference in topographic index at any resolution (1, 5, 10, and 20 m). Even though the correlation values of the regression lines were all negative, the absolute values were very low and thus it can be concluded that differences in topographic indices explained only a small part of the correlations between soil moisture at different locations. Concerning the relation between correlation and difference in texture, again a decrease in correlation was found with increasing difference. However, this relationship was less significant than for distance or differences in height in most cases. Finally, the more the temporally averaged wetness at 2 sensors differed, the lower the correlation between the sensors.

The correlation length varies with the function fitted to the data. A gradient-expansion algorithm was used to compute a non-linear least squares fit of a simple exponential function to the observed relation between correlation and difference in terrain features:

$$r(h) = \exp\left(-\frac{|h|}{L_E}\right) \tag{3.9}$$

with $|h|$ the absolute difference in distance or difference in terrain characteristics. Since the calculated correlation coefficients are very high and a simple exponential function only reaches the correlation length at 1/e, the correlation length L_E takes rather high values (see table 3.9).

To avoid spurious long range correlations from having any impact on the propagation of assimilated information, several authors [Hamill *et al.*, 2001; Keppenne and Rienecker, 2002] used a localized correlation structure. To determine a correlation length for localization, a compactly supported second order autoregressive (SOAR) function [Gaspari and Cohn, 1999] was fitted to the observed relation between correlation and difference in terrain features:

$$r(h) = \left(1 + \frac{|h|}{L_S}\right) \cdot \exp\left(-\frac{|h|}{L_S}\right) \tag{3.10}$$

The parameter L_S represents the correlation length and is indicated on the plots in figure 3.26, 3.27, 3.28 and 3.29. The correlation length L_S is the distance where the correlation reduces to $2/e$ and where the defined correlation function shows a point of inflection. Note that $|h|$ and L_S are only real 'length' [m] measures when the evolution of the correlation with distance is studied (first column of the figures). The units of these measures vary with the difference considered. For the OPE^3 field, the correlation length decreased with depth and this parameter was clearly higher in winter than in summer for all depths. In the summer, the field was covered with corn, which may contribute to the variability and cause a shorter correlation length. The correlation length for the whole OPE^3 field decreased with depth for L_S from approximately 600 to 50 m for distance (see

also table 3.9), 8 to 2 m for height, 4 to 1 % for clay, 13 to 1 % for sand, 10 to 1 % for silt, 7 to 2 vol% for soil moisture and from 5 to 1 for topindices difference at 1 m resolution. Furthermore, the correlation length was notably larger when considering the whole OPE^3 field compared to the individual sub-watersheds (data not shown). This may be attributed to the fact that the extent of the data set is limited for the individual sub-watersheds [Western and Blöschl, 1999]. Some care is advised in the interpretation of the correlation lengths, because they depend on the chosen structure of the correlation function. For example, L_S was found to be 2.5 to 3 times smaller than L_E.

3.8.2 Spatial characteristics of soil moisture

Investigation and use of the spatial structure of soil moisture is getting more attention with the increase in the availability of remote sensing data. Knowledge of the spatial structure is important in the development [Montaldo and Albertson, 2003] and calibration of 3D hydrological models and in data assimilation studies in general. Refsgaard [2000], Grayson *et al.* [2002] and Wealands *et al.* [2005] reviewed some types of pattern information, their characterization and use with distributed modeling, and pattern comparison methods in the light of calibration and model testing.

From the figures 3.13, 3.14 and 3.15 with soil moisture time series, a large spread in soil moisture over the OPE^3 field at each moment in time can be observed. Using the field averaged soil moisture as a best guess for a soil moisture estimate at a single point is consequently prone to a large uncertainty. This observed spread could be used as an indication for the ensemble spread if the field averaged soil moisture was to be modeled by a probabilistic run.

Descriptive spatial statistics

In figure 3.30 the average spatial soil moisture (m) at different depths over the whole OPE^3 over time is simply calculated as the average of the observations of all operational sensors at a specific time step. Additionally, the standard deviation $stdv$ and coefficient of variation CV=$(stdv/m) \cdot 100\%$ at each time step is plotted. In the calculation of the spatial statistics, a maximum of only one missing sensor was allowed. The evolution of the spatial mean and standard deviation in time gives an indication of the temporal evolution of the spatial soil moisture distribution. Several spatial controls operating at different scales are influencing soil moisture patterns [Wilson *et al.*, 2004], with as most important controls terrain, soil texture, precipitation patterns and vegetation.

Western *et al.* [2004] found that the spatial soil moisture variance was related to changes in spatial averaged soil moisture in the top 30 cm of the soil profile for different sites they studied. However, for each site the relation was different due to different soil properties and meteorological conditions.

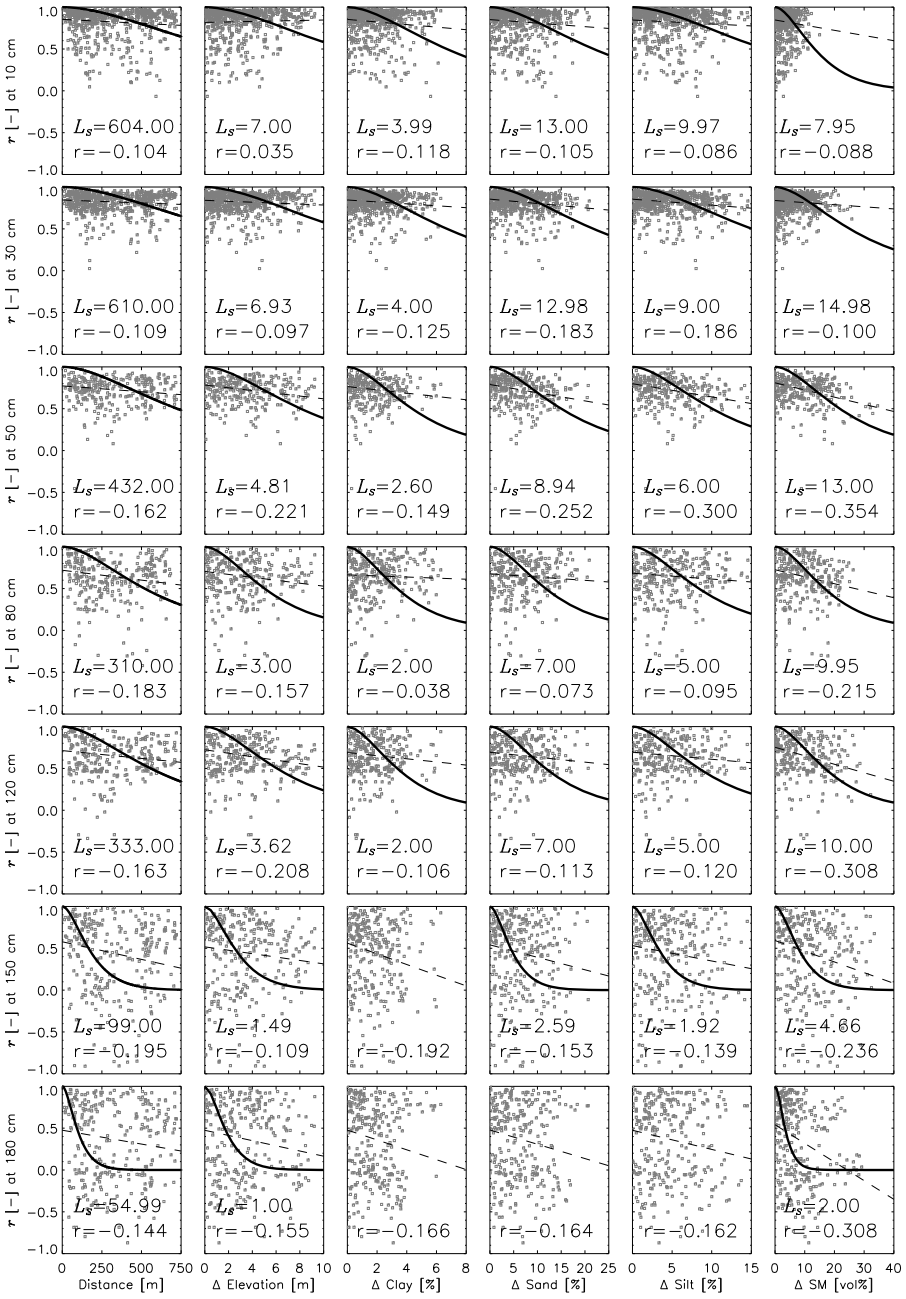


Figure 3.26: Relation between cross-correlation (r) between time series and distance between sensors, difference in height (Δ Elevation), texture difference (Δ Clay/Sand/Silt) and difference in wetness (Δ SM) between sensors in the whole OPE^3 field over the period of May 1, 2001 through April 30, 2002. The correlation length L_S is for the fitted SOAR (—), where fitting was possible. The correlation coefficient (r) is for the linear regression (- -). Note that texture is sampled from the top 9 cm for each plot.

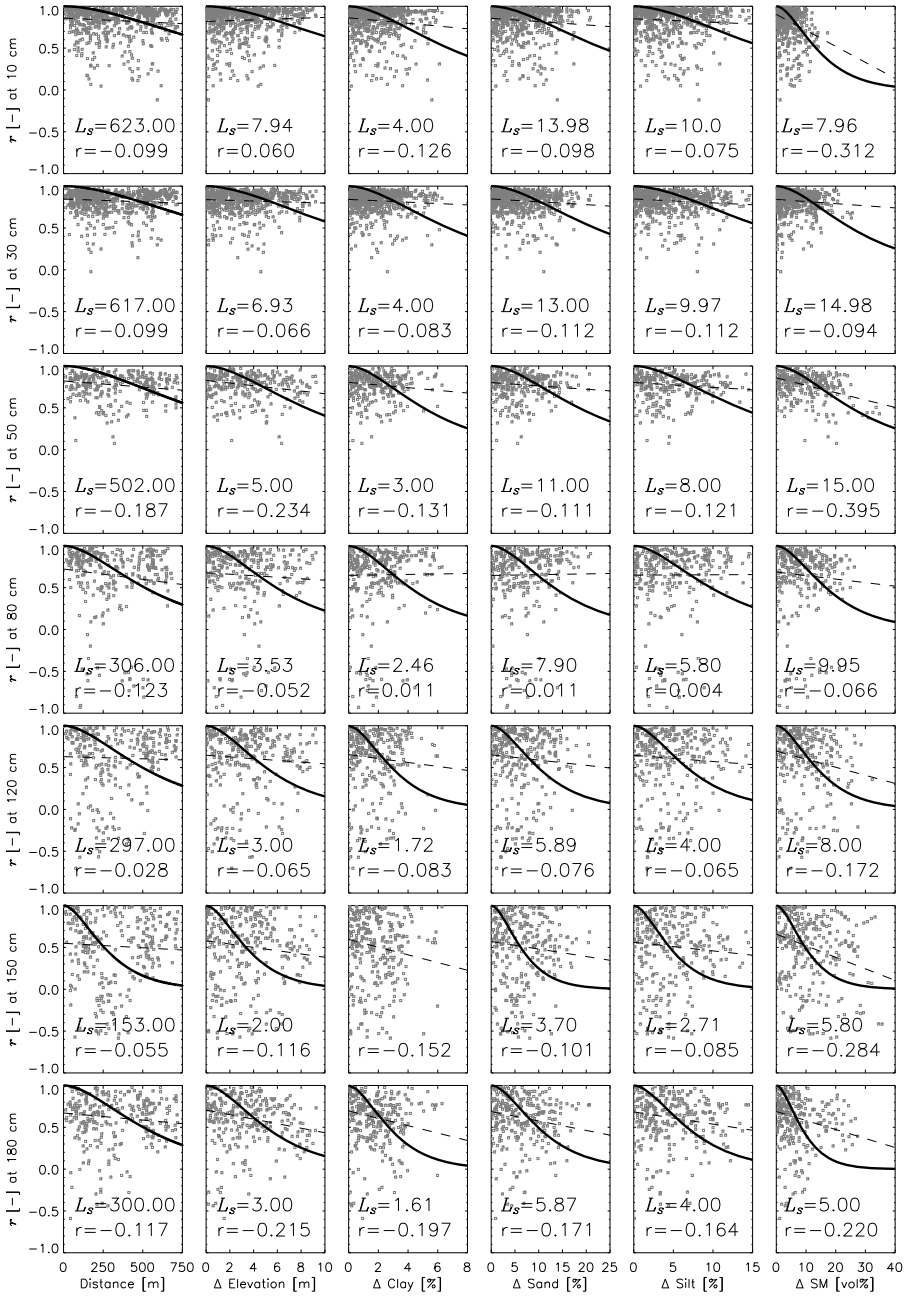


Figure 3.27: Same as in figure 3.26, but for the summer only.

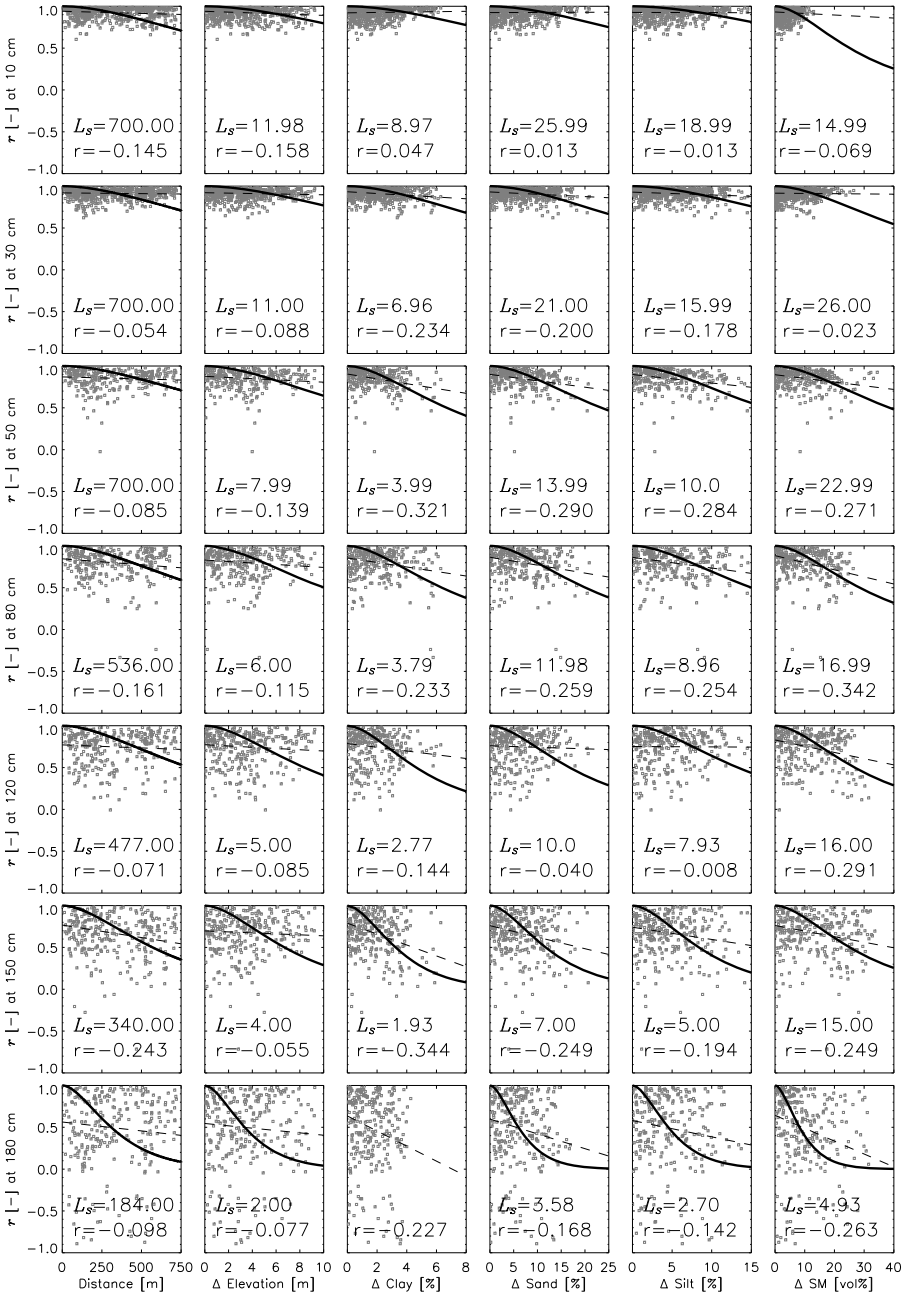


Figure 3.28: Same as in figure 3.26, but for the winter only.

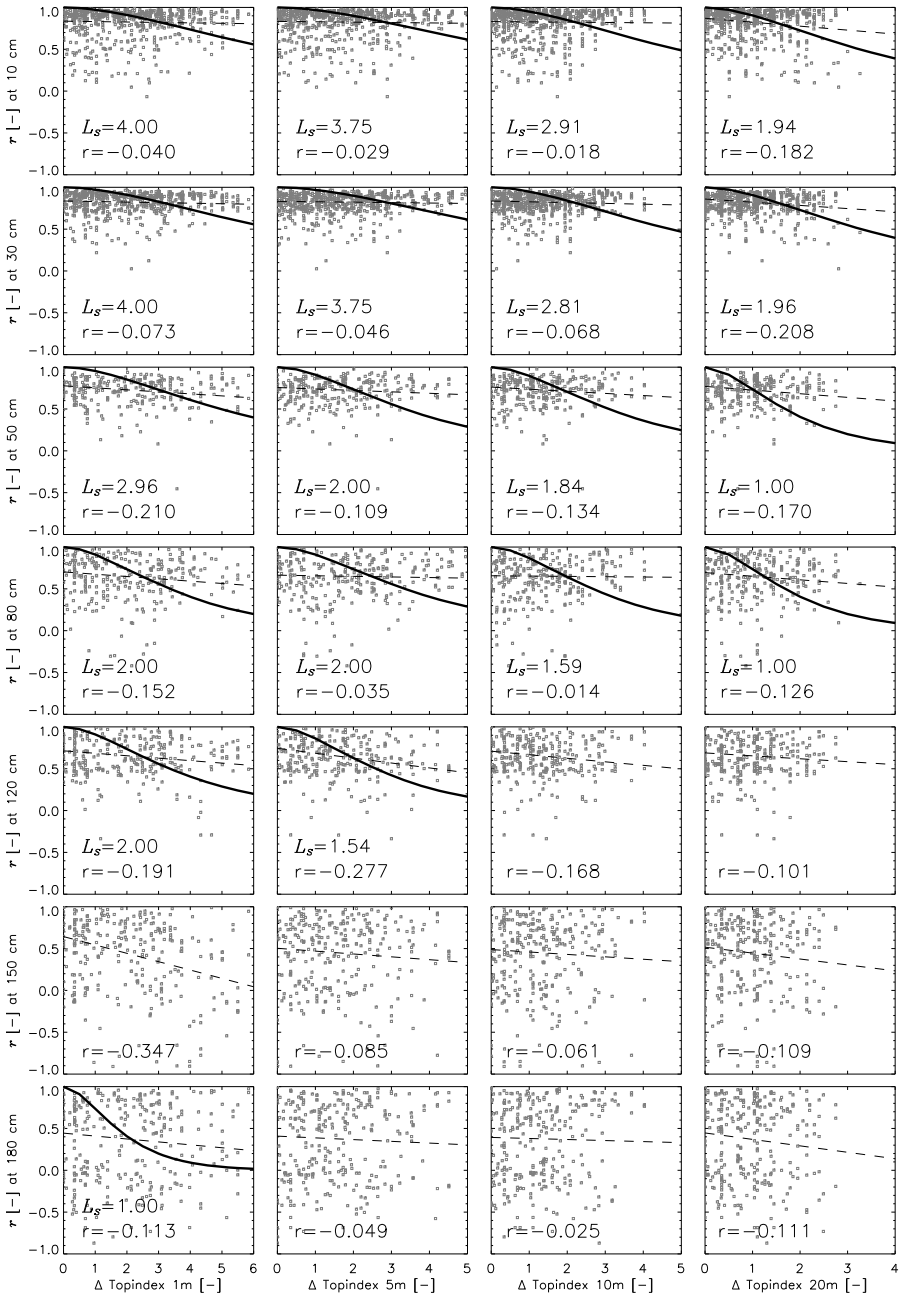


Figure 3.29: Relation between the cross-correlation (r) between time series of soil moisture and difference in topographic index at different resolutions between sensors in the whole OPE^3 field over the period of May 1, 2001 through April 30, 2002 with the legend as in figure 3.26.

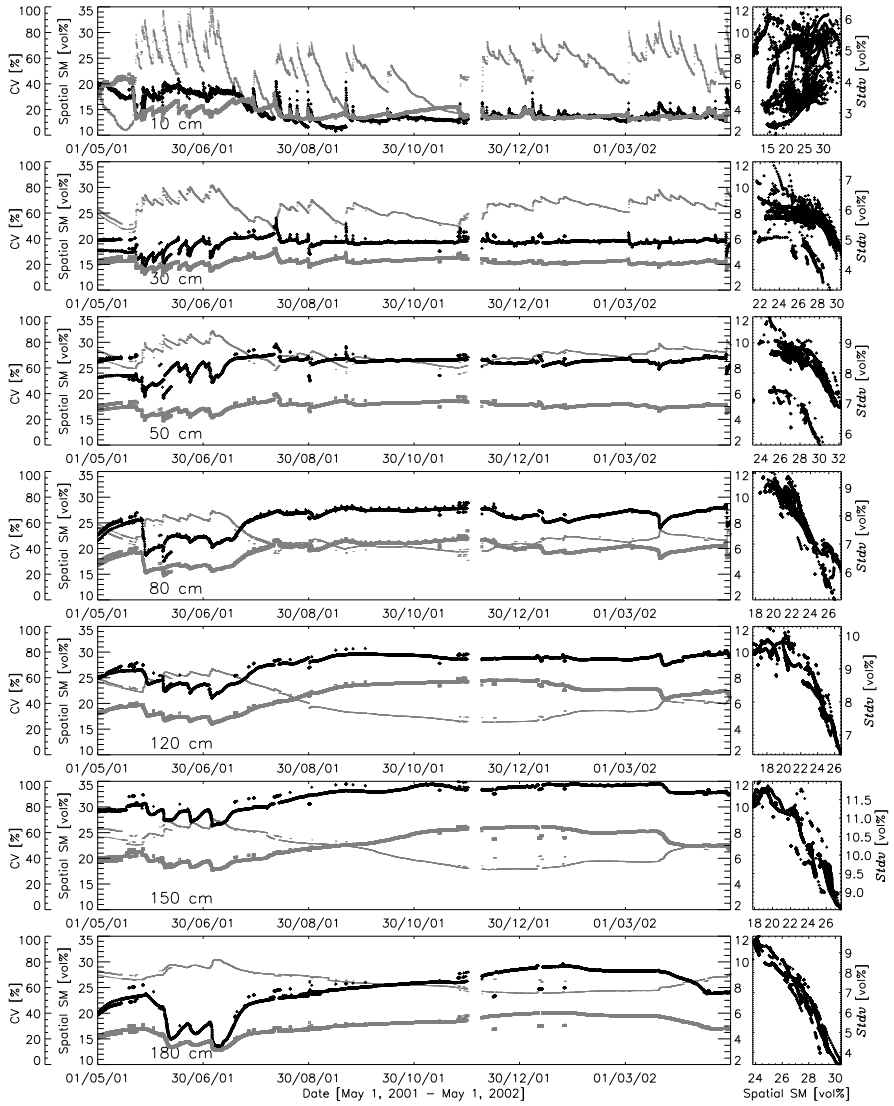


Figure 3.30: Spatially averaged soil moisture SM (thin gray, left vertical scale), standard deviation *stdv* (black +, right vertical scale) and coefficient of variation CV (gray □, extra left vertical scale) at different depths for all sensors in the OPE^3 field during the period May 1, 2001 through April 30, 2002. The righthand side figures plot the standard deviation, observed at the different time steps, in function of the spatial average soil moisture at the corresponding time steps.

For the upper layer at 10 cm depth, the time series in figure 3.30 show peaks in standard deviation when the soil moisture is higher. The scatter plots contain a number of dry-out events, with varying spatial variability for similar average soil moisture. In regression analyses using time series, autocorrelation of the residuals is a problem, and leads to an upward bias in estimates of the statistical significance of coefficient estimates, so that no regression line, nor any indication of the significance is included. During the wet summer of 2001, the variability reached maximal values and did not further increase with increasing soil moisture. It is interesting to see that during late summer and beginning autumn of 2001, where repeated periods of dry outs occurred, the coefficient of variation started to increase, once the soil was drying out, although the standard deviation did not alter much. This can be attributed to the fact that due to spatial differences in evapotranspiration, the dry out rate will not be spatially constant, causing that the standard deviation on the soil moisture decreases less rapidly than its mean value, resulting in an increase in coefficient of variation. Also the spatial variability may remain high due to local ponding and runoff caused by the intensive rainfall events. During the summer of 2001, an opposite behavior can be seen. This can be attributed to the fully grown, spatially uniform canopy and an overall high evapotranspiration rate, which reduces the spatial differences in dry out. Through simulations, Teuling and Troch [2005] found that both soil and vegetation controls can either create or destroy spatial variance and this depends on whether or not the soil dries below a critical soil moisture content, which depends on texture and precipitation.

At and below 50 cm, the spatial standard deviation and coefficient of variation decreased with increasing areal soil moisture, clearly in the wet summer period of 2001: the higher the soil moisture, the lower the variability. Less variability was found under wet conditions for the lower depths, because these layers were kept moist by the groundwater table, limiting the variability in wet circumstances. Dry out caused an increase in variability for the deeper layers.

At 30 cm depth, the standard deviation showed a behavior with features of the upper layer at 10 cm as well as of the lower layers at and below 50 cm: the standard deviation sometimes increased and sometimes decreased with increasing spatially averaged soil moisture. Clearly, soil moisture in the upper layers (corresponding to the rooting depth) showed a different spatial variation than the lower layers. Further, the spatial standard deviation and the coefficient of variation were larger in deeper layers than in upper layers at all times. The limited coefficients of variation near the surface explain the high values for the cross-correlation for the surface layers in section 3.8.1. The high coefficients of variation for deeper layers explain the lower values for the cross-correlation and reveal that accurate quantification of the subsurface soil water dynamics is a difficulty in the OPE^3 field.

From the scatter plots of areal soil moisture versus standard deviation, it becomes clear that for shallow depths in particular a possible relationship would be time dependent or, in other words, for a given spatial average of soil moisture, the variation or spatial distribution differed at different times. For deeper depths, one

may usually expect the same spatial variation each time a same spatial average is observed.

Grayson *et al.* [1997] defended the hypothesis that soil moisture patterns switch between two 'preferred states', being the wet state dominated by lateral flow and the dry state dominated by vertical fluxes. Therefore, they studied the temporal changes in soil moisture to see the different response in a dry summer and wet winter time. From spatial soil moisture data it was found that the pattern was more random in dry situations and that more connection was found for wet situations. This is in correspondence with the finding that the standard deviation decreased with increasing spatial averaged soil moisture in the OPE^3 field for deeper layers, possibly due to lateral redistribution of water, with the deeper layers probably more sensitive for the influence of lateral flow than the upper layers.

It should be noted that an increase in standard deviation as discussed above may not necessarily be linked with a more random pattern and less connectivity. Connectivity is a feature that cannot be captured by standard geostatistical approaches [Western *et al.*, 2001], but it may reveal interesting information on preferred flow paths and lateral flow movement. Further research on this topic should be performed to better understand this phenomenon.

At 10 cm depth, an increase in the standard deviation with wetness was due to precipitation, which is possible as long as no saturation is reached. Note that similar findings were reported by several authors [Robinson and Dean, 1993; Famiglietti *et al.*, 1998; Western *et al.*, 1998], whereas a decrease in spatial variability with increasing soil moisture was found by Famiglietti *et al.* [1999] and Buttafuoco *et al.* [2005]. The wet state at the surface is thus dominated by a vertical flux (local control) that can be quite random in space: vegetation causes a variable interception and the presence of rows in the corn field creates an additional variability between row and inter-row areas [Van Wesenbeek and Kachanoski, 1988]. Under dry conditions at the surface, there is a spatially almost constant boundary condition determined by atmospheric forcing, which causes evapotranspiration. Because the evapotranspiration rate decreases with decreasing moisture content, the moisture content of wetter soils will decrease faster than in dryer soils. As a result, the spatial variability will decrease during dry out. These observations confirm recent studies indicating that surface soil moisture variability shows a concave pattern with mean soil moisture [Albertson and Montaldo, 2003; Ryu and Famiglietti, 2005]

General spatial statistics of soil moisture are summarized in table 3.8. Time averages of spatial statistics were taken over the time period of 1 year (averaging over the time series of 1 year), winter time and summer time. All calculations were performed on hourly averaged data. Wilson *et al.* [2004] found that the temporal variation in soil moisture was 5 times the spatial variation. This was not true in the OPE^3 field, where the temporal (see figure 3.16) and spatial (see table 3.8) variation were of the same magnitude.

Table 3.8: Overview of the time-space (m) averaged soil moisture [vol%] and the temporally averaged spatial standard deviation ($stdv$) [vol%] for different periods for all sensors in the individual sub-watersheds, as well as in the complete OPE^3 field (All) and for the spatial averages of the 4 sub-watersheds within the complete field (4sub).

Depth [cm]	A m	$stdv$	B m	$stdv$	C m	$stdv$	D m	$stdv$	All m	$stdv$	4sub m	$stdv$
1 year (May 1, 2001 - April 30, 2002)												
10	22.3	3.2	20.6	3.1	25	3.7	23.3	3.5	22.5	3.9	22.7	1.7
30	28.4	2.4	22.1	4.9	30.9	6.9	26.5	4.1	26	5.7	26.9	3.2
50	32.9	2.8	21.7	6.6	34.5	12	28.1	5.7	27.6	8.4	29.1	5.0
80	26.4	8	17.8	6.8	19.3	10.2	25.4	6	22	8.3	22.3	3.7
120	20.1	6.4	17.9	7.6	22.8	10.3	21.3	9.9	20.1	9.1	20.5	2.4
150	20.2	6.8	22.7	8.9	27.5	11.6	19.5	12	21.7	10.9	22.3	3.6
180	25.1	3.7	28.5	5.5	30.5	5.7	22.5	9.4	26.1	7.8	26.5	3.2
summer period only (May 1, 2001 - October 30, 2001)												
10	21.5	3.1	19.6	3.4	24.0	4.2	22.9	4.0	21.7	4.2	22.0	1.9
30	28.1	2.5	21.9	4.8	30.6	7.1	26.4	4.1	25.7	5.6	26.5	3.1
50	33.0	2.9	21.9	6.5	34.8	11.8	28.2	5.5	27.7	8.3	29.2	5.0
80	27.1	6.4	18.6	6.7	20.4	10.3	25.4	6.2	22.6	8.0	23.0	3.5
120	21.8	5.5	21.2	7.1	24.8	9.0	22.0	10.0	22.1	8.7	22.4	2.1
150	22.8	6.0	26.4	7.2	31.1	9.0	19.6	11.9	23.9	10.3	24.7	4.4
180	27.0	2.3	30.0	3.7	31.5	3.7	23.4	8.7	27.4	6.8	27.8	3.2
winter period only (October 31, 2001 - April 30, 2002)												
10	23.1	3.4	21.5	2.9	25.8	3.2	23.8	3.0	23.3	3.5	23.5	1.6
30	28.8	2.3	22.4	5.1	31.1	6.8	26.6	4.1	26.3	5.7	27.2	3.2
50	32.9	2.8	21.4	6.6	34.2	12.2	27.9	5.9	27.4	8.5	29.1	5.0
80	25.6	9.8	17.0	6.9	18.4	10.2	25.4	5.8	21.3	8.7	21.6	3.9
120	18.4	7.3	14.4	8.0	21.0	11.4	20.7	9.7	18.1	9.5	18.6	2.7
150	17.5	7.6	18.8	10.6	24.3	14.0	19.3	12.0	19.4	11.5	19.9	2.8
180	23.1	5.1	26.9	7.4	29.5	7.5	21.6	10.1	24.8	8.9	25.2	3.2

Spatial correlation and change with time

The spatial correlation $r(h)$ at a lag or separation distance h at 1 time instant is calculated by the formula

$$r(h) = \frac{E[(z_{\mathrm{x}} - \overline{z_{\mathrm{x}}})(z_{\mathrm{x}+h} - \overline{z_{\mathrm{x}+h}})]}{\sqrt{E[(z_{\mathrm{x}} - \overline{z_{\mathrm{x}}})^2]E[(z_{\mathrm{x}+h} - \overline{z_{\mathrm{x}+h}})^2]}} \tag{3.11}$$

with $\overline{z_{\mathrm{x}}}$ the space averaged value of z_{x}, $E[(z_{\mathrm{x}} - \overline{z_{\mathrm{x}}})^2]$ the variance and $C(h) = E[(z_{\mathrm{x}} - \overline{z_{\mathrm{x}}})(z_{\mathrm{x}+h} - \overline{z_{\mathrm{x}+h}})]$ the covariance between variable $z_{\mathrm{x}+h}$ and the same variable at a distance (in this case: in space) h away from x:

$$E[(z_{\mathrm{x}} - \overline{z_{\mathrm{x}}})(z_{\mathrm{x}+h} - \overline{z_{\mathrm{x}+h}})] = \frac{1}{n(h)} \sum_{j=1}^{n(h)} (z_{\mathrm{x}_j} - \overline{z_{\mathrm{x}}})(z_{\mathrm{x}_j+h} - \overline{z_{\mathrm{x}+h}}) \tag{3.12}$$

$$= \frac{1}{n(h)} \sum_{j=1}^{n(h)} z_{\mathrm{x}_j} z_{\mathrm{x}_j+h} - \overline{z_{\mathrm{x}} z_{\mathrm{x}+h}} \tag{3.13}$$

with $\overline{z_{\mathrm{x}}}$ and $\overline{z_{\mathrm{x}+h}}$ the means of the head and tail values of the data pairs separated by a vector h, and $n(h)$ the number of pairs. Possibly changing means with varying h can be taken into account (non-ergodicity) [Goovaerts, 1997]. The covariance only exists under the condition of strict stationarity. It is statistically not correct to calculate spatial correlation immediately by formula (3.11) for the sensors in the OPE^3 field, as the number of observations at one time instant is not sufficient and the condition of stationarity cannot be guaranteed.

Experimental evidence showed that for the OPE^3 field, the correlation and covariance are highly dependent on the direction of the pairs. That is why for this practical case one is better off to use variograms [Van Meirvenne *et al.*, 2005], which are not as much restricted by the condition of stationarity (existence of a constant mean and finite variance is not required) as spatial correlation functions and where only the difference between values is important and not the order in which they are processed. A possibly remaining problem is that semivariances are built under the assumption of random variables, while soil moisture is often found to be spatially organized [Western *et al.*, 1999] and characterized by the feature of connectivity. The semivariance $\gamma(h)$ at a lag h is expressed by

$$\gamma(h) = \frac{1}{2n(h)} \sum_{j=1}^{n(h)} (z_{\mathrm{x}_j} - z_{\mathrm{x}_j+h})^2 \tag{3.14}$$

with $n(h)$ the number of couples separated by an equal distance of lag h. In this study, only omni-directional variograms were considered, assuming that the spatial variability is identical in all directions, even though this might be a violation of reality. For a stationary random function the covariance $C(h)$, the correlation $r(h)$ and the semivariance are related by:

$$\gamma(h) = C(0) - C(h) \tag{3.15}$$

$$r(h) = \frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)} \tag{3.16}$$

The higher the semivariance, the lower the correlation.

At each observation time, variograms were calculated. It is well known that soil moisture has a high variability and variograms are often hard to fit. For example, Williams *et al.* [2003] found variograms with a pure nugget effect and Skøien *et al.* [2003] concluded that soil moisture measured in small catchments (0.1 to 1.5 km^2) was non-stationary and close to fractal over the extent sampled. For soil moisture in the OPE^3 field, the variogram fitting was sometimes found to be difficult, when not enough pairs of observations were available at certain lags.

After some tests for several models at some individual time steps, it was decided to fit exponential variogram models to the experimental semivariance values, as in Western *et al.* [1998]. The exponential variogram is given by:

$$\gamma(h) = C_l \left[1 - \exp\left(\frac{-3\ h}{a}\right)\right] \tag{3.17}$$

The parameters range, a and sill, C_l were sought by least squares minimization of the error between the observed and the fitted variogram. The sill C_l is the level at which the variogram flattens out. The presence of a sill indicates stationarity of the soil moisture distribution [Goovaerts, 1997]. The nugget, which represents small scale variability and measurement errors, was not included to reduce the computational load for automated fitting, since it was found experimentally that this value was very low. Inclusion of the nugget $C(0)$ in the expression for the exponential variogram is simply done by addition of this term to Eq. (3.17). For an exponential variogram, as defined in Eq. (3.17), the numerical value of the range a is 3 times the correlation length of a simple exponentially decreasing correlation function. Consequently, the spatial correlation length does exactly correspond to the value of b in another common expression for the exponential variogram:

$$\gamma(h) = C_l \left[1 - \exp\left(\frac{-h}{b}\right)\right] \tag{3.18}$$

The correlation length is a measure of the spatial continuity of the variable and the range is the distance beyond which the correlation between points is minimal. The correlation length is reported to be scale dependent and typically found to increase with the catchment scale [Western and Blöschl, 1999]. Furthermore, semivariances can only be calculated for lag distances smaller than $1/2 \times$ the maximal dimension of the area under study. For comparison, Vinnikov *et al.* [1996] found a spatial correlation scale of 400 to 800 km for soil moisture up to 1 m deep in the middle latitudes of the northern hemisphere. Western and Blöschl [1999] discussed the fact that statistical properties appearing in data, such as the apparent variance and apparent correlation length, are different from their true values, because bias is introduced by the measurement scale, which is defined by the spacing, extent and support of the data sample. Furthermore, with scale there are important

changes in processes controlling the moisture patterns, which explains differences in correlation lengths. Petrone *et al.* [2004] tried to determine the optimal spatial scale and sampling density to find representative values of the statistical structure, variance and spatial autocorrelation, since these values were found to be dependent on the sampling resolution and grid size. It is worth to recall this when these statistics are used, e.g. in a data assimilation scheme.

Western *et al.* [2004] studied the variation in variogram parameters for 6 to 8 time instants over 2 years in different catchments. They found that spatial variance and thus also the sill and, to a lesser extent, the correlation length were related to the mean soil moisture. Their study of variograms was based on many points in space and probably yielding more accurate estimates of parameters than obtained for the OPE^3 soil moisture measurements. However, the strength of the application to the OPE^3 field measurements lies in the close to continuous time series of measurements, which will reveal more clearly the evolution of the variogram parameters in time. Inclusion of the dynamic behavior of the spatial structure during system identification is expected to improve modeling and data assimilation results, as was illustrated in meteorology by e.g. Bouttier [1994].

Experimental results
Omni-directional variograms were calculated with a lag intervals of 10 and 20 m, allowing for some averaging, as for smaller lag intervals the amount of pairs contributing to each lag would be insufficient to even consider the use of variograms [Van Meirvenne, 2003]. Fitting of exponential variograms to the observed semi-variances, gave optimal values for the parameter range, a, and sill, C_l, at each time step. The resulting parameters were prone to errors due to the limited sample in space, the large lag resolution and the choice of the predefined fixed variogram model, which may not be suited at each time step. Since less averaging over pairs was possible at a 10 m lag resolution, mainly the parameters fitted at 20 m lag resolution will be discussed. The evolution in time of the range and the sill at different depths is shown in figure 3.31. These parameters were clearly varying with the soil moisture content in time. At some time instants, the parameters values did not show a smooth evolution, due to missing sensors.

From the time series of geostatistical information in figure 3.31 for different soil depths of the whole OPE^3 field over the complete time series of 1 year, following conclusions can be made. At 10 cm depth, the range of the fitted variogram was temporally very variable and evolved proportionally to moisture: the range mostly decreased during dry-out. This was not so clear from the scatter plots, because these include different sequences of wetting and drying. The range was around 200 m and during the wet summer of 2001, the range reached 300 m. During dry periods, vertical fluxes, i.e. evapotranspiration, regulated the soil moisture values in the upper soil layer, and the range decreased. Under very dry conditions, the variability in space was very low and the shapes of the variogram suggested a pure random nugget effect. During wet periods with a vertical flux of precipitation, the range increased, which means that the precipitation (throughfall) values were higher correlated than the evapotranspiration in space. The range values were

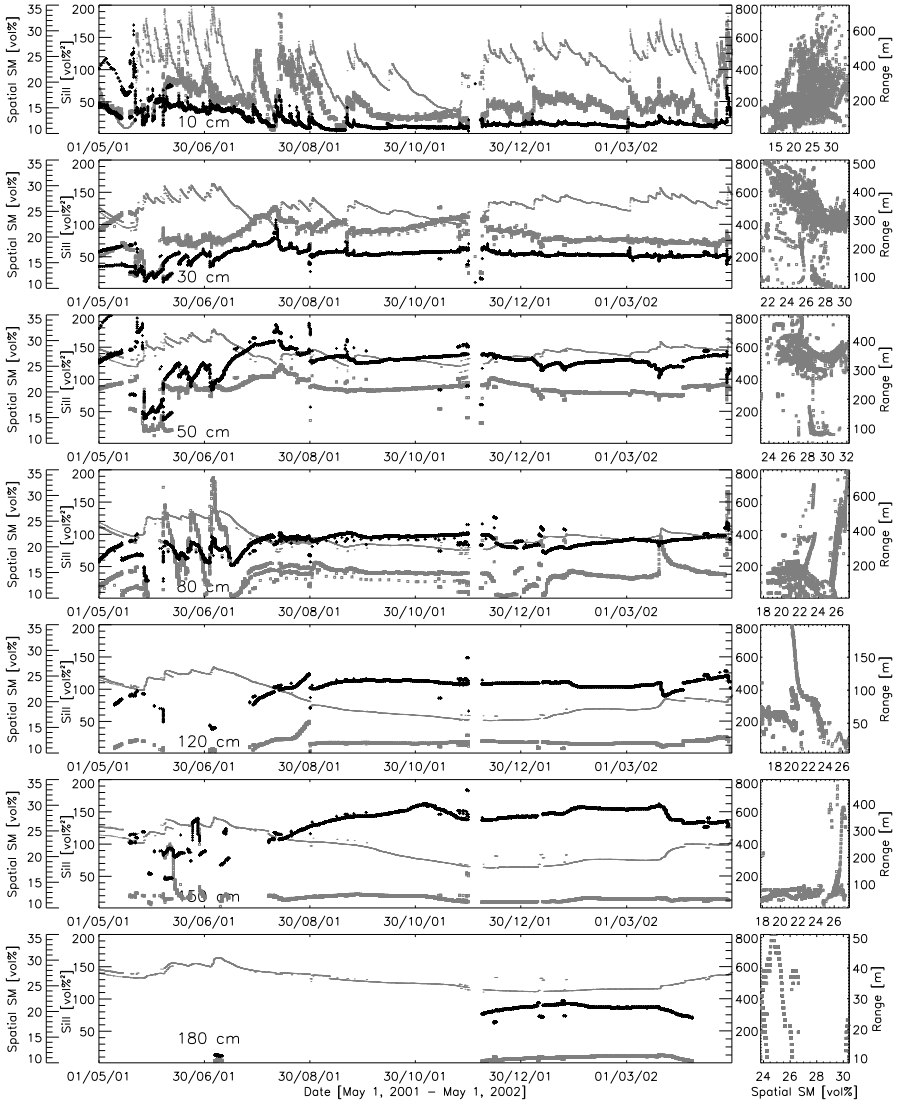


Figure 3.31: Time series of variogram parameters (range and sill in gray □ and black +, with right and left vertical scale, respectively) and spatially averaged soil moisture SM (thin gray, extra left vertical scale). Some spurious points are caused by missing sensors, which affects the estimation of the geostatistical parameters. At deeper layers, variogram fitting was not always possible, which results in missing parameters. The righthand side figures plot the range, observed at the different time steps, in function of the spatial average soil moisture at the corresponding time steps.

highest at 30 cm and 50 cm. This is because at these depths, water from the top layer is redistributed and these layers are not or little affected by processes like preferential flow and dry-out in deeper layers. The range reached about 300 m at 30 cm and 50 cm and from 80 cm depth onwards, the range values decreased with depth. This can be attributed to the complex hydrogeology with different processes occurring at different locations.

Evidently, the evolution of the sill is directly linked with the evolution of the spatial variance (see spatial standard deviation in previous paragraph). Comparison of the spatial variance with the fitted sills revealed that indeed the sill and the variance showed a very similar temporal behavior, which also means that the variogram fitting has been performed with enough reliability to capture the temporal changes in parameters.

Except for the geostatistical parameters at 30 cm depth, the range and the sill evolved similarly in time: at a given depth, a higher variability was observed together with a larger range. At deeper depths with high sills, low ranges were found, independent of the time. Of course for all findings, it was assumed that the best variogram model remained always exponential, which may not always be the case.

Western *et al.* [1998] found that the geostatistical structure for the top layer soil moisture evolved seasonally within 1 year. High sills (15-25 vol%^{2}) and low correlation lengths (30-50 m) were observed during a wet winter period and smaller sills (5-15 vol%^{2}) and longer correlation lengths (50-60 m) were found in a dry summer time in the 10.5 ha Tarrawarra catchment [Western and Grayson, 1998]. Based on more different catchments, Western *et al.* [2004] concluded that typical correlation lengths were between 30 and 60 m and that the relation of the correlation length with the soil moisture content was dependent on the catchment. Further, for the same Tarrawarra site, an increase in correlation length with wetness was found during another study year. Their results demonstrated that the processes controlling the spatial patterns of soil moisture change with the soil moisture status in time and in space.

Time integrated spatial correlation

For simplicity, several state estimation techniques assume a time-invariable static correlation structure of the variables, causing that the radius of influence is kept constant through time. To obtain integrated sill and range values over the entire time period studied, these parameters could be calculated in several ways. Since the fitting of instantaneous variograms may not always be optimal, a simple averaging of the ranges or derived correlation lengths is not advised. Therefore, pooled stratified (in time) variograms were proposed as an alternative to calculate the average correlation length.

Semivariogram values were calculated at each hourly time step (stratification in

time) and all obtained variograms during the study period were pooled to end up with multiple semivariance values at each lag. Then, an exponential variogram model was fitted to the observed values. Because of the large spread in semivariance values at larger lags, the fitting for these pooled variograms was limited to the values for the first 200 m only. An example of such a pooled stratified variogram with averaged semivariance values for each lag over 1 year is shown in figure 3.32 for soil moisture at 10 cm depth. Pooled stratified (in time) variograms at 10 and 80 cm depth for each month are shown in figure 3.33 and 3.34, respectively. In table 3.9, geostatistical parameters for pooled stratified (in time) variograms over 1 year for the whole OPE^3 field are summarized for variograms calculated with a lag interval 20 m (similar for other lag intervals). Note that for an exponential variogram, the numerical value of the range is three times the correlation length. It was impossible to find a proper fit for the soil moisture structure at 180 cm. Integrated over time, the sill was lower (15-40 vol%^{2}) for upper layers than for lower layers (around 100 vol%^{2}) and the correlation length reached maximal values around 70 m at 30 cm and decreased with depth to values less than 20 m. Notice that the temporally averaged spatial standard deviations (and hence the variances) in table 3.8 show the same tendency as the sill values, which indicates that the variogram fitting was quite accurate. The spatial correlation lengths obtained for the simple exponential and SOAR fitting are also given in table 3.9 and evolved similarly over depth as the correlation lengths deduced from variogram fitting.

Table 3.9: Geostatistical parameters describing the spatial structure of observed soil moisture over the whole OPE^3 field. The range and sill were derived from fitting an exponential variogram to the pooled (over 1 year) stratified (in time) semivariance values. The correlation length in the upper block of the table is calculated as the range/3. The correlation lengths L_E and L_S are the parameters of the simple exponential correlation function and the SOAR, respectively, that were fitted to the cross-correlations (integrated over time) versus distance between sensors.

Depth	10 cm	30 cm	50 cm	80 cm	120 cm	150 cm	180 cm
Sill [vol%^{2}]	17	37	72	67	93	119	/
Range [m]	143	201	139	53	53	37	/
Correlation length [m]	48	67	46	18	18	12	/
Correlation length L_E [m]	1977	1988	1162	748	808	280	162
Correlation length L_S [m]	604	610	432	310	333	99	55

3.8.3 Correlation lengths

In subsection 3.8.1, exponential functions and SOAR functions were fitted to the relation between cross-correlation and distance between sensors. In subsection 3.8.2, the spatial correlation was determined for each time step through variogram fitting. Both approaches yield values for a correlation length which is related to distance. These measures are of primary importance for data assimilation applications, but in the literature the meaning and underlying principles for both approaches have not been thoroughly studied. A correct interpretation of correlation values and correlation lengths (beyond which it is not needed to account for correlations and thus computational savings may be introduced) may offer a great contribution to the improvement and optimization of data assimilation techniques.

Cross-correlation between time series of a pair of sensors gives an integrated view of the correlation over a predefined time interval. The correlation length found by fitting an exponentially decreasing function, depends on the expression of this function. Since the cross-correlation values between the time sensors is very high (when it rains, there is an increase in soil moisture for all sensors within the OPE^3 field), it will not easily decrease to the value of $1/e$ or $2/e$ for pairs of sensors (mainly in the top layer) situated in the area of the OPE^3 field. Consequently relatively long correlation lengths are expected before correlation coefficients reduce to $2/e$ and even longer values could be expected before the correlation decreases to $1/e$.

Variograms are calculated for single time steps and an evolution of the range in time can be observed. The range is found where 95% of asymptotical value of the sill is reached, which corresponds to an approximation of the correlation r to an asymptotical minimal (zero) value. A rule of thumb for practical purposes is to assume that the practical range is 3 $\times$ the spatial correlation length (cfr. the discussion around Eq. (3.17)). An advance of using variograms to calculate the correlation length is the availability of time-varying information, which can be included through time-dependent localization in Kalman filter approaches in particular, while only a time averaged value of correlation length is needed for more simple methods.

The time-integrated correlation lengths found for soil moisture based on pooled variograms were within the range of values found by Western *et al.* [2004]. However, these correlation lengths were much smaller than the values derived from a SOAR fitting to the cross-correlations as function of distance as in section 3.8.1. It has already been discussed that the correlation values and hence the correlation lenths were relatively high for sensors within the OPE^3 field, mainly because the sensors are imposed to similar meteorological forcings. Probably, this correlation length refers to processes that act on a larger scale than the one that is considered by the variogram fitting. Vinnikov *et al.* [1999] suggested that the soil moisture variability consisted of two components, one of which was related to large scale atmospheric forcing and another to small-scale land surface variability and hydrological processes. Based on spatial autocorrelation functions, they proposed

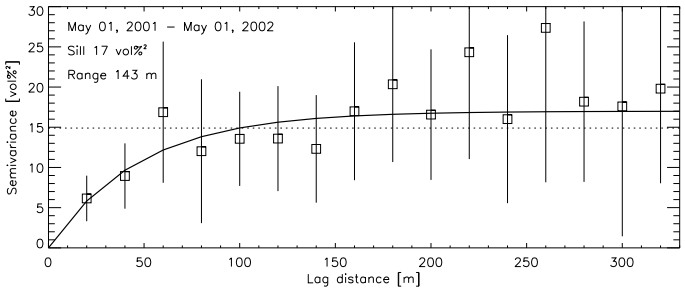


Figure 3.32: Observed (□) pooled stratified (in time) variogram at 10 cm over 1 year for the whole OPE^3 field with indication of 1 standard deviation. The full line represents the best fitted exponential model. The dashed line shows the time averaged spatial variance.

methods for optimal design of surface networks for soil moisture observations.

3.8.4 Point vs. spatially averaged values

In this section, data for individual sensors are compared to the spatial OPE^3 mean soil moisture. For each sensor the temporal average and standard deviation of the differences from the spatial average over the period of May 1, 2001 through April 30, 2002 were calculated to search for representative sites. Further, the issue of up-scaling is discussed.

Time or rank stable sites

Grayson and Western [1998] proposed a method to determine areal estimates of soil moisture based on point measurements in catchments with significant relief to overcome the mismatch in scale between the classical point measurements in field experiments and the areal estimates from remote sensing and modeling. One could use geostatistical techniques to interpolate point measurements, but given the limited correlation lengths observed for soil moisture, this is not a practical alternative. Another method is to use wetness indices based on terrain (and at best texture) data. However, it has been discussed that they are based on a number of assumptions that make them practically useless. Grayson and Western [1998] investigated the existence of certain locations that consistently show the mean areal soil moisture behavior, irrespective of the overall wetness and the pattern of soil moisture. Such locations were named Catchment Averaged Soil Moisture Monitoring (CASMM) sites. This approach is based on the concept of time stability, introduced by Vachaud *et al.* [1985]. Several authors have explored this approach to analyze soil moisture patterns in different catchments [Kachanoski and de Jong, 1988; Gómez-Plaza *et al.*, 2000; Van Pelt and Wierenga, 2001; Martínez-Fernández

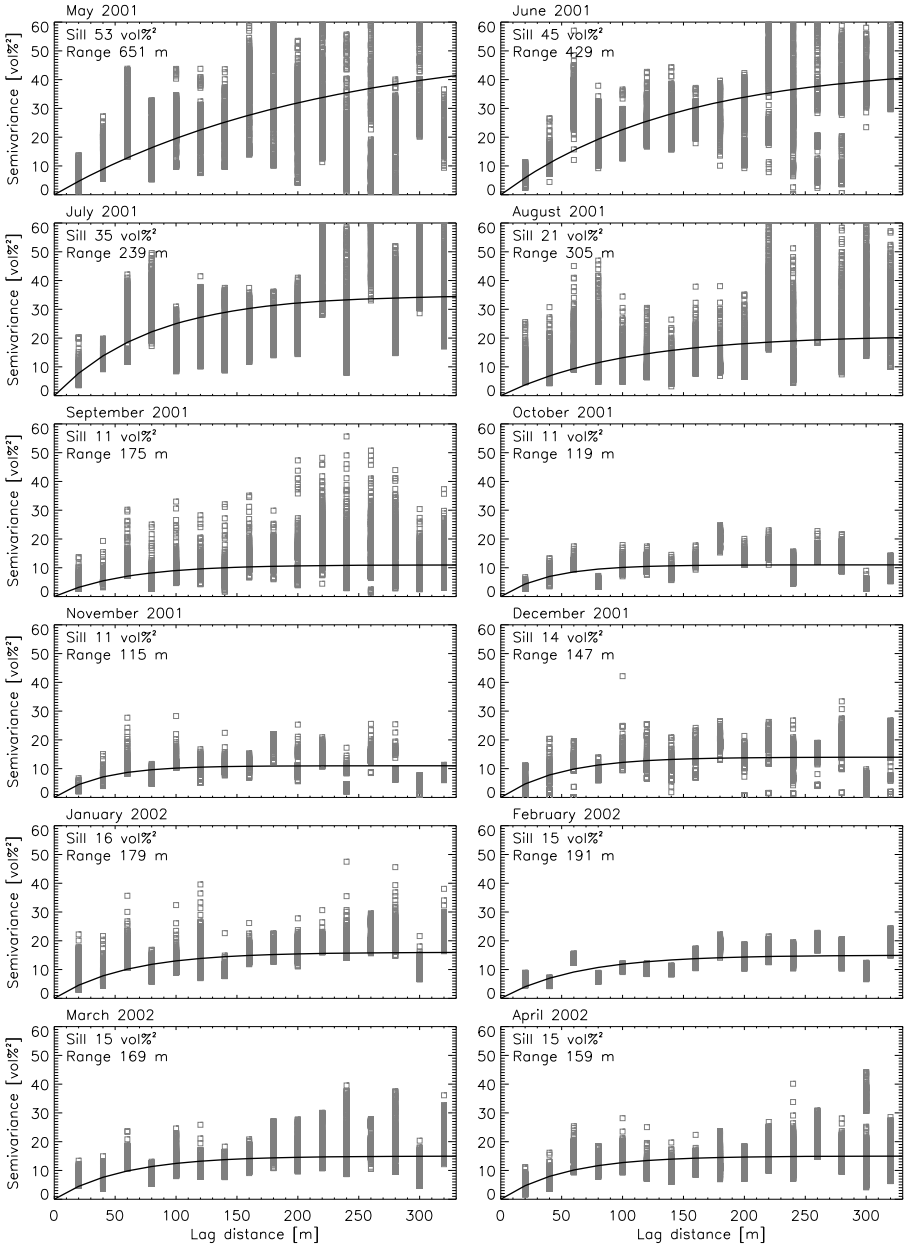


Figure 3.33: Observed (gray □) pooled stratified (in time) variograms at 10 cm for each month in the period May 1, 2001 through April 30, 2002 for the whole OPE^3 field. The line represents the best fitted exponential variogram model.

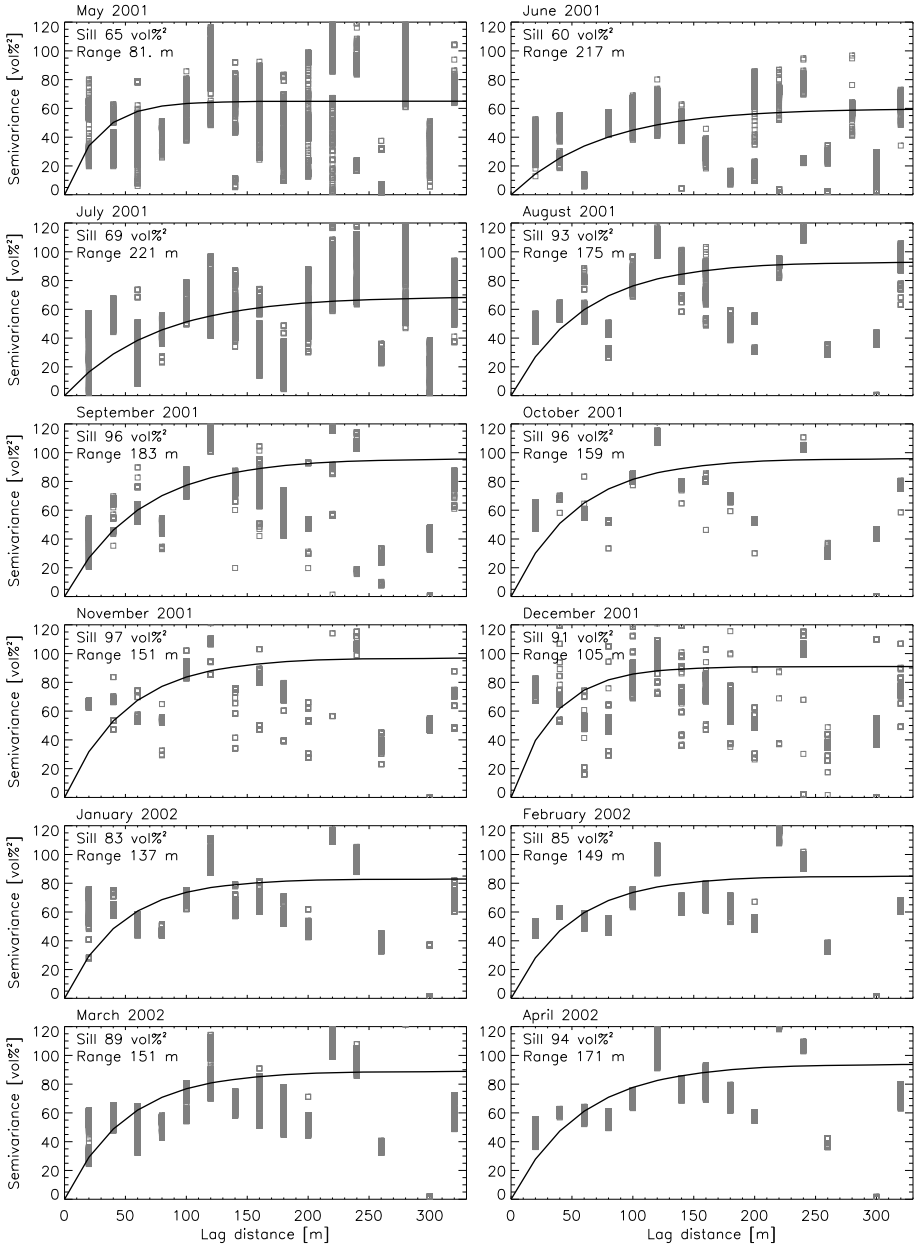


Figure 3.34: Observed (gray □) pooled stratified (in time) variograms at 80 cm for each month in the period May 1, 2001 through April 30, 2002 for the whole OPE^3 field. The line represents the best fitted exponential variogram model.

and Ceballos, 2003, 2005; Cosh *et al.*, 2004; Jacobs *et al.*, 2004]. Chen [2006] argued that it was more appropriate to refer to rank stability instead of time stability. Time stability would suggest a similarity in temporal changes, while the focus of rank stability is rather on temporally persistent soil moisture patterns, which are expected to be influenced by soil texture, vegetation and topography.

Except for their use for validation of and comparison to remote sensing data, time stable sites are interesting for an unexplored application in state estimation. Assimilation of data coming from a CASMM site can be expected to be much more efficient than assimilation of data from a site that does not show the same behavior as the rest of the catchment. If information from a CASMM site is propagated to other sites, it is very likely that the other sites are properly updated, while assimilation of strange information will probably lead to poor state estimation results. The idea of CASMM sites can be used to distinguish between more and less interesting data points for assimilation.

With $SM_{j,i}$ the soil moisture for a single point or sensor j at each time step i, and $\overline{SM}_i$ the spatially averaged soil moisture at each time step i, the relative difference $d_{j,i}$ of the soil moisture content for a sensor is calculated by:

$$d_{j,i} = \frac{SM_{j,i} - \overline{SM}_i}{\overline{SM}_i} \tag{3.19}$$

The time averaged relative difference, $< d_j >$, gives an indication of how much the sensor deviates from the spatial mean during the entire time period under consideration. However, the mean relative difference does not provide an unbiased estimator in case of noise on the $\overline{SM}_i$ signal. Through the above definition of relative difference, differences from the spatially averaged soil moisture have a lower weight in $< d_j >$, in case the spatial averaged soil moisture is high. Note that there is a multiplicative relationship between the spatially averaged soil moisture value and the soil moisture value at a point:

$$SM_{j,i} = d_{j,i}\overline{SM}_i + \overline{SM}_i, \quad \text{and} \quad \overline{SM}_i = \frac{SM_{j,i}}{d_{j,i} + 1} \tag{3.20}$$

If the temporal evolution of $SM_{j,i}$ is to be estimated based on $\overline{SM}_i$ with $d_{j,i}$ replaced by its time average, $< d_j >$, then this expression indicates that for those sensors with higher soil moisture values than the spatial average (i.e. positive $< d_j >$), there is more temporal variability than for the average soil moisture. If $\overline{SM}_i$ is to be estimated based on $SM_{j,i}$, then the amplitude of $\overline{SM}_i$ will be smaller than that of $SM_{j,i}$, if $< d_j >$ is positive. The opposite is true for negative $< d_j >$.

Since previous studies, nor the data at hand show that this assumption is based on physical reality, also the absolute difference $SM_{j,i} - \overline{SM}_i$ was studied in this work. Notice that here 'absolute' contrasts to the 'relative' difference and does not mean positive differences. An advantage is that through this approach there is an additional relationship between the spatially averaged soil moisture value

and the soil moisture value at a sensor or point. Consequently, the temporal variability of the spatial mean is assumed to be identical to that for individual point measurements, which is a more realistic assumption. Furthermore, since no weighting by $1/\overline{SM}_i$ is applied in the temporal averaging, there is less distortion, which allows for an easier interpretation of the statistics.

The standard deviation in the (relative/absolute) difference $d_{j,i}$ for each sensor is of primary importance as a small standard deviation refers to a consistent deviation and thus such kind of sensor shows a similar temporal evolution in soil moisture as the areal soil moisture. With n the number of time steps, the standard deviation for sensor j is given by:

$$stdv(d_j) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{N}(d_{j,i} - < d_j >)} \tag{3.21}$$

Independent of the kind of difference calculated, a representative site can be identified as one for which $< d_j >$ is close to 0. A small temporal standard deviation in the (relative/absolute) difference $d_{j,i}$ for a sensor j, $stdv(d_j)$, implies that this sensor shows a similar temporal evolution in soil moisture as the areal soil moisture. A sensor with this feature is called time or rank stable and can be used as representative for the areal soil moisture, if the offset $< d_j >$ between the areal soil moisture and the soil moisture at the sensor is known. The most attractive representative site would be one for which both $< d_j >$ and $stdv(d_j)$ are close to 0, or in other words for which the root mean square difference between the time series of point measurements and the spatial mean values is minimal.

To study the stability of the soil moisture field, several authors use the Spearman rank coefficient [Vachaud *et al.*, 1985] or the correlation coefficient as defined by Chen *et al.* [1997]. These methods will not be further explored here.

Experimental results

While originally the CASMM site approach was developed for soil moisture in the root zone, it was explored for deeper layers also. In table 2 in appendix A the experimental time averaged results are summarized for absolute differences between areal soil moisture of the sub-watersheds and the point values in these sub-watersheds. Figure 3.35 shows the time averaged absolute differences and their standard deviation for each sensor relative to the whole OPE^3 field, while figure 3.36 shows similar plots for relative differences. Also the spatially averaged soil moisture for sets of sensors in each of the 4 sub-watersheds were compared to the averaged soil moisture over the whole OPE^3 field. Since at some time steps some sensors had missing data (which results in a varying spatial mean), the mean differences and their standard deviation, were calculated only for those time steps for which a maximum of 1 sensor was missing.

There are only slight differences in the ranking of the sensors dependent of the

use of a absolute difference or a relative difference and of course the sign of the values remains equal. The higher the averaged difference for a sensor, the more its measurements deviate from the spatial average. The higher the standard deviation, the less consistent the measurements deviate from the average. So sensors with a low standard deviation in deviations follow more the general trend of the spatial average, but their measurements can be consistently higher/lower than the spatial average. Also, one can expect a higher correlation between areal soil moisture and a time series from a sensor with a low standard deviation in deviations.

One could select a sensor with a low $stdv(d_j)$ as being representative for the field average plus or minus a constant value, that is preferably close to 0. For example, sensors in probes DH1 and DH2 were very close to the areal mean at 10 cm depth and were characterized by differences of limited variance. At 10 cm depth, sensor BM3 had the smallest variability (lowest $stdv(d_j)$) in absolute and relative differences, but differs on average more from the areal mean. Sensor DM3 showed less variability in relative differences at this depth, but this could not be retrieved for the absolute differences. The sensors with the lowest $stdv(d_j)$ for both types of differences at deeper depths were AM1 at 30 cm, DM3 and BL2 at 50 cm, AH1 and BL4 at 80 cm, BM3 at 120 cm, AM1 at 150 cm and AM2 at 180 cm depth. It was difficult to find a probe that had a small $stdv(d_j)$ over all depths for the studied time period, but one could propose probes AM1 and BL4 for the deeper profiles and DH1 for the shallow profiles. When considering all depths, examples of probes that are not representative for the spatial mean behavior, as they had high values for $stdv(d_j)$, are e.g. probes DH4, BL1 and BL2. The deviant behavior of these probes could be attributed to the fact that they are situated at the highest elevations of the watershed. More stable probes are located along the hillslope. However, note that DH4 (near the top) had a very low $< d_j >$ value in the 2 upper layers. Individual probes in field C were in general also less representative than other probes, which supports the fact that this field has a different subsurface hydrology. Sensors with lowest $< d_j >$ values for both types of differences were BH4 at 10 cm, DH1 at 30 cm, DM2 at 50 cm, BL4 at 80 cm, BL4 and BM4 at 120 cm, BL3 and BL4 at 150 cm and AM2 and BL3 at 180 cm. One could propose probes BH1 and AH2 for shallow profiles and probes AM2 and DL3 for deep profiles as probes with overall the smallest $< d_j >$ values at most depths. Soil moisture values at probe CM1 and DM2 were most different from the spatial mean soil moisture: CM1 was too wet, whereas DM2 was too dry to be representative.

As spatial averages of some sensors are more likely to follow the total spatial averaged means than individual sensors, because some variability is smoothed out, the spatially averaged soil moisture for the 4 sub-watersheds were compared to the total averaged soil moisture over the whole OPE^3 field. The error when taking a spatial average of one of the sub-watersheds as representative for the whole OPE^3 field is of course less than taking only one of the sensors. The temporal evolution of the complete soil moisture profile was in general much better represented by the averaged soil moisture over any sub-field than by a single probe, i.e. $stdv(d_j)$ was

always very low. However, some sub-fields were considerably wetter/drier than the complete field at some depths, such as field B at 50 cm. From table 3.10, it can be concluded that field A is most rank stable (smallest $stdv(d_j)$) at 50, 80, 120 and 150 cm depth. At 10 cm depth, field B showed the highest stability in time, and at 30 and 180 cm depth, field C could be selected as most representative for the mean spatial soil moisture.

From the time series of the differences for the individual sensors (not shown), it is clear that even for the sensors with the lowest $< d_j >$ values some temporal variability as in soil moisture remains. As illustrated in figure 3.37 for the sensors in field B, none of the sensors really holds a fixed position relative to the spatial mean (over field B only in this case) during the studied time period of 1 year, if the sensors are sorted following their difference from the mean for every time step. It is possible to find sensors that hold a fixed position over shorter time intervals. Some sensors take over each others position relative to the spatial average, definitely for the upper layers in periods with a high frequency of precipitation events and in summer time. Also for the case of subsets of point data, it is clear that none of the sub-watersheds keeps a fixed position relative to the total spatial mean (figure 3.38) and that the deviations still show temporal variations in soil moisture.

Analysis of the temporal cross-correlation matrices (figures 3.21, 3.22 and 3.23) and the ranking of the mean cross-correlation of a sensor with all other sensors at the same depth (Figure 3.25) revealed that for larger values of $stdv(d_j)$, a lower mean temporal cross-correlation with all other sensors can be expected, mainly for the upper layers. Since in the OPE^3 field the spatial correlation length was found to be longer for higher mean cross-correlation, this also indicates that smaller $stdv(d_j)$ corresponded to larger correlation lengths. For example, for the top 10 cm, sensors DH4, DH3, AH1 and CM1 show a very low average cross-correlation in Figure 3.25 and a high $stdv(d_j)$ in Figure 3.36. However, this relation between low average cross-correlations and high $stdv(d_j)$ is not apparent for all layers. This link between the spatial correlation length and $stdv(d_j)$ might be very interesting for spatially distributed modeling with horizontally linked grid cells. If observations were assimilated at points with a large soil moisture correlation length (which might occur where $stdv(d_j)$ is low), then the influence of assimilating these data is expected to have more impact than soil moisture assimilation at a point where the soil moisture is only marginally correlated with the surrounding points.

A more detailed analysis of the representativeness of the different sensors in the OPE^3 field is reported by De Lannoy *et al.* [2006d].

How representative is the SCAN site?

Figure 3.39 shows time series for both the spatially averaged OPE^3 soil moisture and the single point time series of the SCAN site during the overlapping period of available data. The soil moisture at the SCAN site was consistently lower than in the OPE^3 field. Since there was a wide soil moisture spread at the point

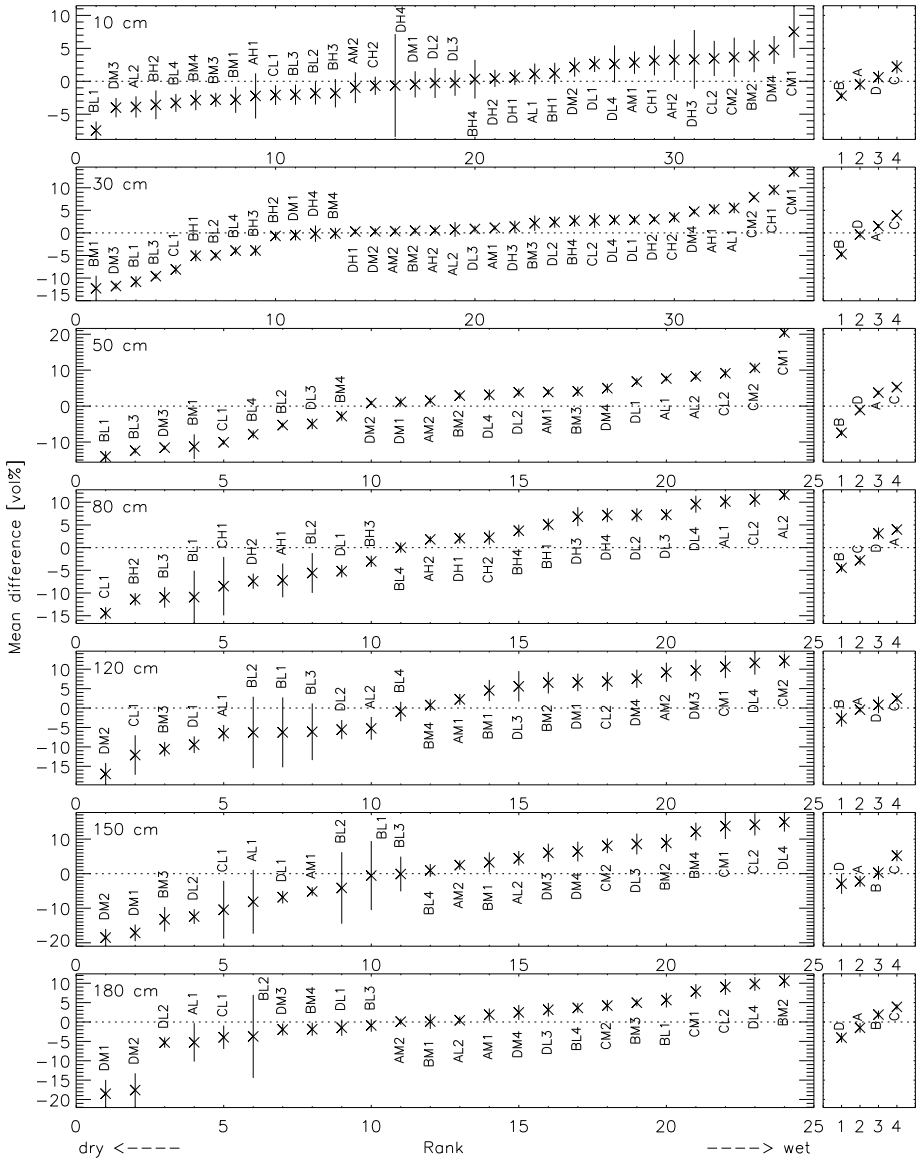


Figure 3.35: Average and standard deviation in absolute differences between the areal wetness and point soil moisture measurements for all sensors in the whole OPE^3 field. The sensors are sorted following their average deviation from the areal wetness. The average deviations are calculated by almost (but not exact) the same number of samples over 1 year for every sensor, i.e. about 8550. For the C-sensors there are less samples, i.e. about 7900, for AH2 there are only about 5400 samples.

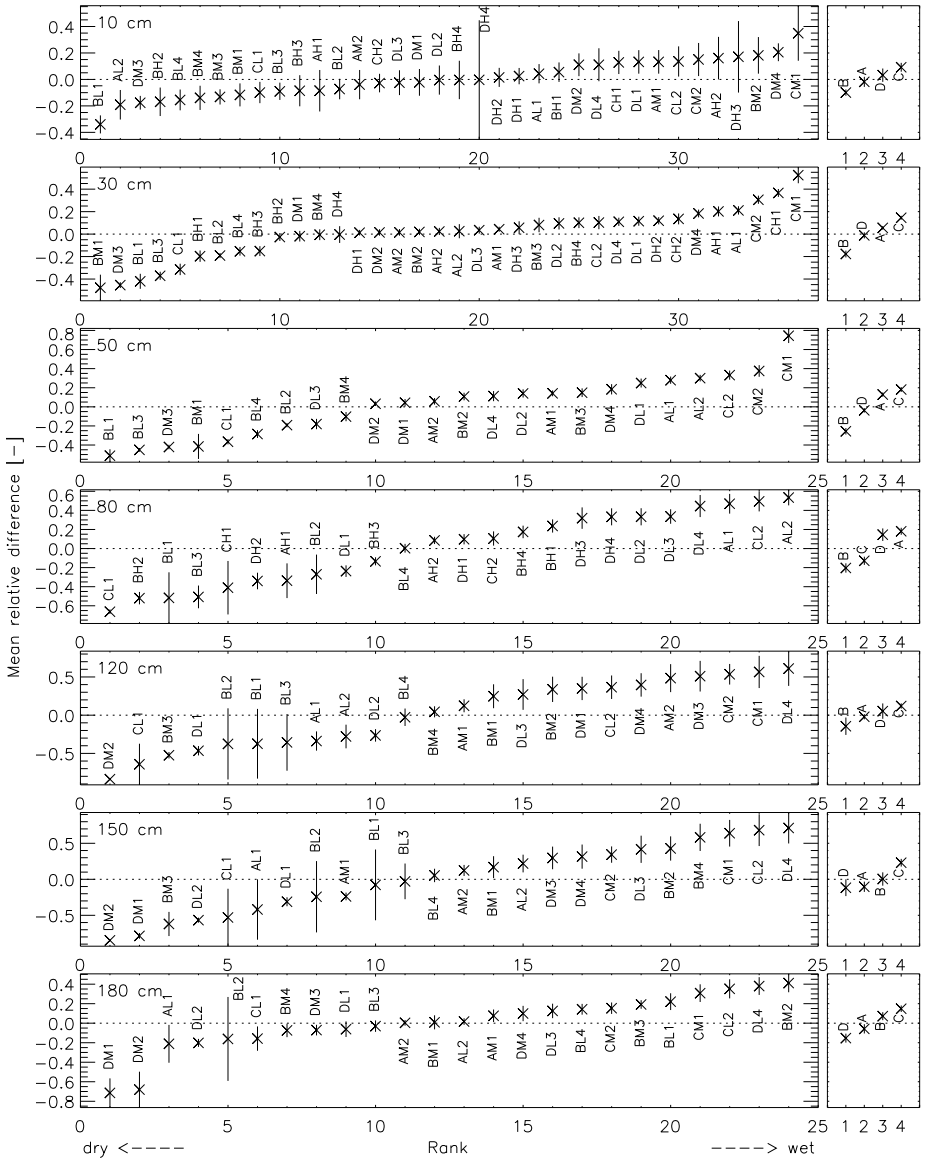


Figure 3.36: Average and standard deviation in relative differences from the areal wetness for all sensors (left) and for the individual fields (right) within the whole OPE^3 field. The sensors are sorted following their average deviation from the areal wetness. The average deviations are calculated by almost (but not exact) the same number of samples over 1 year for every sensor, i.e. about 8550. For the C-sensors there are less samples, i.e. about 7900, for AH2 there are only about 5400 samples.

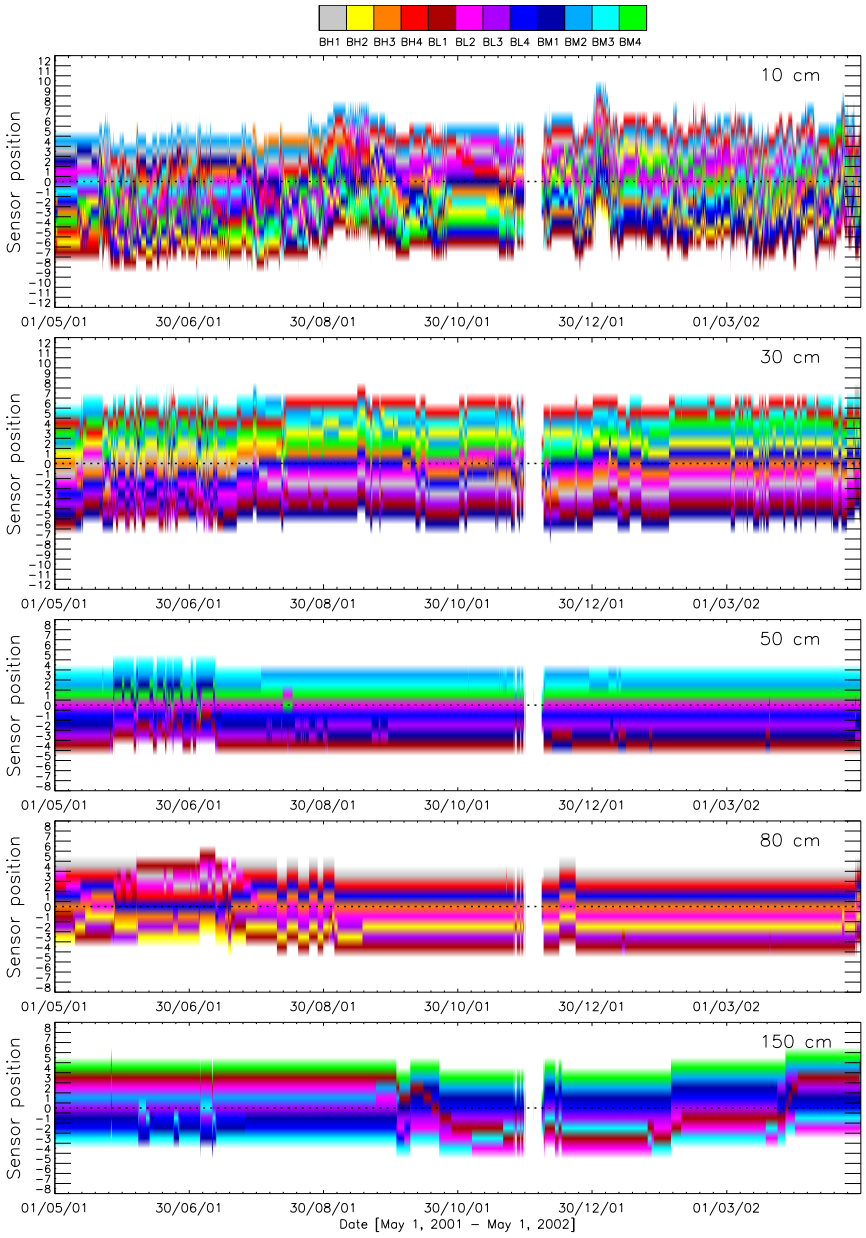


Figure 3.37: Relative position of the point measured soil moisture of sensors in field B to the spatial average soil moisture. The sensor best corresponding to the spatial average is positioned at 0. The higher the sensor position, the higher the soil moisture.

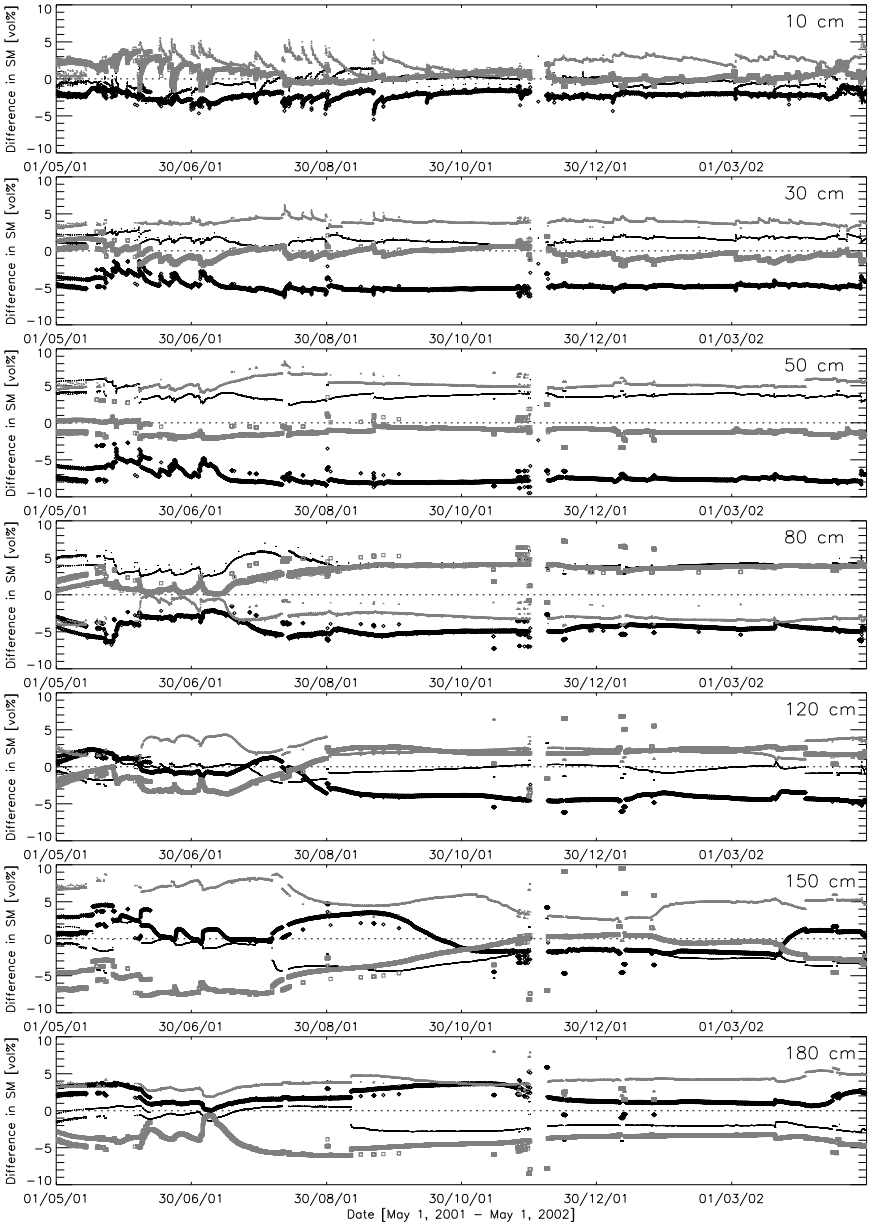


Figure 3.38: Absolute differences from the areal wetness for the 4 individual subcatchments (thin black + : field A, thick black ◇: field B, thin gray △: field C, thick gray □: field D) relative to the whole OPE^3 field over the period May 1, 2001 through April 30, 2002.

Table 3.10: Overview of how spatial averages of the 4 sub-watersheds behave with respect to the average the entire OPE^3 site, including the mean (m) [vol%] of the absolute differences and the standard deviation ($stdv$) [vol%].

sub-catchment		10 cm	30 cm	50 cm	80 cm	120 cm	150 cm	180 cm
A	m	-0.46	1.50	3.73	3.99	-0.42	-2.21	-1.49
	$stdv$	0.88	0.47	0.56	0.69	0.81	1.57	1.23
B	m	-2.19	-4.72	-7.43	-4.45	-2.67	0.25	1.90
	$stdv$	0.48	0.65	0.85	0.85	2.14	1.99	1.05
C	m	2.20	3.91	5.28	-2.79	2.45	5.21	3.90
	$stdv$	1.02	0.36	0.63	0.92	0.85	1.63	0.84
D	m	0.61	-0.36	-1.09	3.12	0.82	-2.91	-4.05
	$stdv$	0.99	0.82	0.67	1.32	2.15	2.93	1.23

scale in the OPE^3 field (figure 3.13, 3.14 and 3.15), this may cause the averaged OPE^3 data to deviate significantly from the single point SCAN data. Also the measurement depths were not identical. Further, the difference may be partially attributed to the land cover difference: the OPE^3 field is a corn field, the SCAN site is mainly covered by grass. Finally, different sensors (EnviroSCAN versus Hydra probe) are used and there is a difference in the measured soil volume and in the calibration method. SCAN data are widely used over the USA and assumed to be representative for the surrounding area. The SCAN site showed a very similar temporal behavior as the areal mean in the OPE^3 field, but at least a constant term should be added in order to match the spatial mean soil moisture in the OPE^3 field. De Lannoy *et al.* [2006d] discussed how these point data could best be converted to represent the spatial mean OPE^3 soil moisture.

Up-scaling

To estimate spatially averaged soil moisture from point data, e.g. for direct assimilation of the point data into a coarse grid assimilation system, it is beneficial to scale up the point data by an existing relationship, even if these data were collected at a representative site. Some statistical methods were explored, including a time-mean bias correction and cumulative density function (cdf) matching. Also models or filters identified in the frequency and time domain have been studied. The reader is referred to De Lannoy *et al.* [2006d] for the discussion of the results. Crow *et al.* [2005b] demonstrated that up-scaling by using field-scale data only could be enhanced by merging them with model predictions.

3.8.5 Statistics for data assimilation

The observed soil moisture data and their derived statistics provide invaluable information to identify the system, to choose the correct parameters and initial states for models. This inclusion of observational information is data assimilation in large sense. Differences between modeled and observed spatio-temporal statistics are often used as measures of goodness-of-fit [Western *et al.*, 1999].

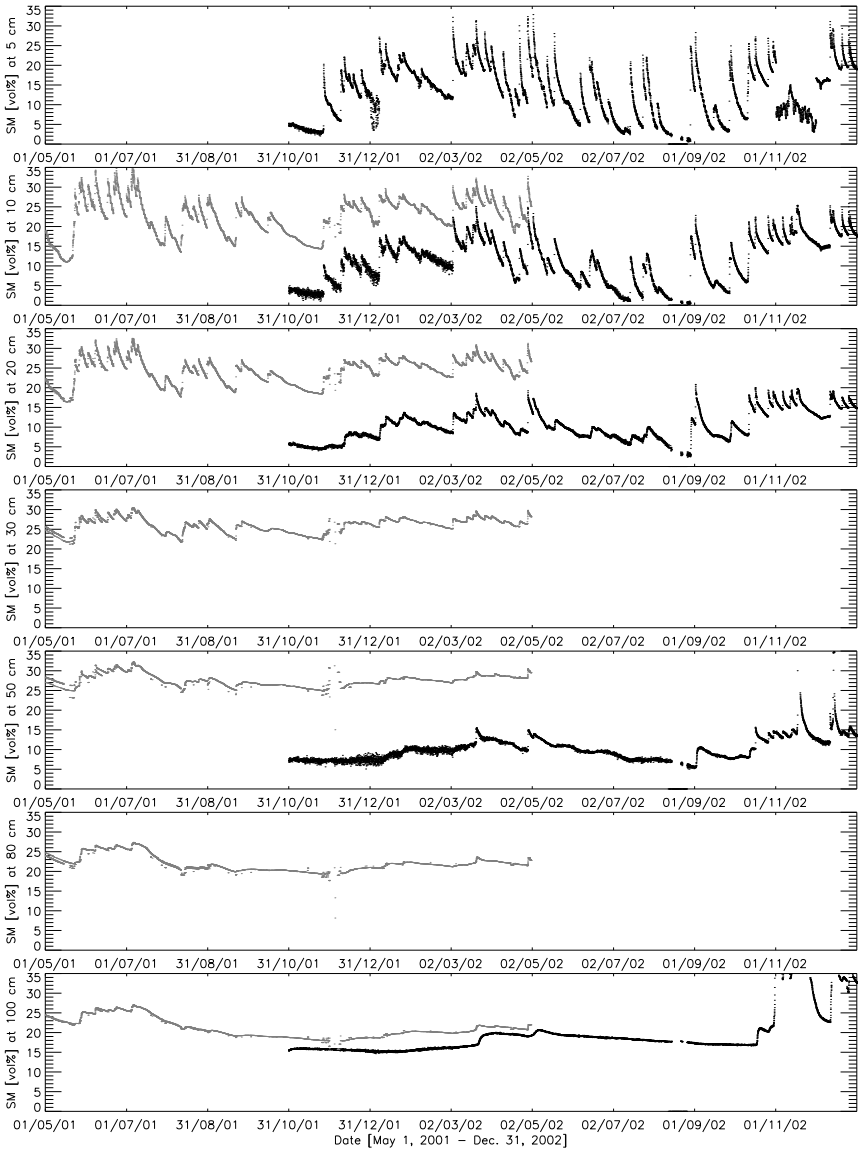


Figure 3.39: Time series of spatially averaged OPE^3 observations (gray) and SCAN (black) observations of soil moisture at different depths. The areal OPE^3 observations at 120 cm depth are used to compare to SCAN data at 100 cm depth.

In the framework of state estimation, information on the spatio-temporal behavior of soil moisture aids to determine the required spatial density of observation points: if the spatial correlation is large, a less dense network of monitoring sites is needed to update a spatially distributed network. For the OPE^3 field, a maximal inter-distance of 70 m (i.e., the correlation length beyond which the correlation becomes smaller than 1/e) is allowed when soil moisture observations at 30 cm are assimilated, whereas when top layer soil moisture is used, this monitoring network should be denser. Furthermore, the assimilation frequency depends on the temporal correlation. For soil moisture measurements within the rooting depth, this correlation reduced to 0.368 (=1/e) after two weeks, whereas for deeper layers, the correlation remained higher than 1/e for several months, indicating that less frequent soil moisture measurements would be needed if the assimilation happens at deeper layers.

It is logical to expect that assimilation of observations from time stable sites is more beneficial for parameter or state estimation than observations from a site which shows a deviating hydrologic behavior. Therefore, good observation points would be found among the CASMM sites. As indicated above, up-scaling of point data may further enhance assimilation results.

3.8.6 Resulting statements

- The temporal variability in soil moisture decreased with depth in the profile, except when there was evidence of sudden dry-out or preferential lateral subsurface flow.
- The temporal autocorrelation increased from about 2 weeks in the rooting zone up to 2 months in the deepest (180 cm) layer.
- The cross-correlation between soil moisture time series was in general quite high in the upper soil layers, where all sensors responded similarly to inputs; the cross-correlation decreased with depth, and at deeper layers, sensors showed different responses due to side effects like lateral flow (nonlocal control).
- The sensors with high spatially averaged cross-correlations typically showed low variance in these cross-correlations with other sensors and were representative for the behavior of most other sensors.
- The cross-correlation decreased with increase in distance, height differences, difference in topindices, difference in texture and difference in averaged soil moisture; through fitting of a simple exponential and a SOAR model, the correlation length was found to decrease with depth and was clearly dependent on the season, on the extent of the studied area (individual sub-watersheds or complete OPE^3 field) and on the chosen model.
- The spatial structure in time averaged soil moisture could not be related with terrain features.

- Soil moisture at individual time steps was often well correlated with terrain features; a multivariate regression between soil moisture and terrain features was used to generate spatial maps of soil moisture, and combined with soil moisture maps generated through kriging.
- The evolution of the spatial variance in time differed for the upper layer and the lower layers; for the upper layers the variance increased with areal soil moisture, due to the local control of the vertical precipitation flux; for the deeper layers the variance was in general higher than for the upper layers and in time less variance was observed for wetter circumstances, probably due to the nonlocal control of lateral flow in the subsurface layers.
- Variogram parameters evolved differently with soil moisture in time for upper and lower layers; the correlation length for pooled variograms over 1 year was about 50 to 70 m in the rooting zone, and 20 m or less in deeper layers; the sill increased with depth.
- Sensors positioned at the highest locations of the OPE^3 field and some sensors in field C (with a different subsurface hydrology) were found to show a behavior that deviated significantly from the areal mean; no general relationship could be found between terrain features and rank stability; all 4 sub-watersheds represented very well, and better than the individual probes, the average soil moisture profile for the complete OPE^3 field.
- Point measurements from the SCAN site showed the same temporal behavior as the whole OPE^3 field, but the values differed from the areal mean by an almost constant value.
- Observation operators were developed to scale up point data to spatial mean soil moisture.

3.9 Summary

In this chapter, the OPE^3 site was described, which will help to parameterize the CLM2.0 in next chapter. Further, an overview of the available data was given and the preparation of the meteorological data as forcings for the CLM was discussed. Finally, the soil moisture data were analyzed for spatial and temporal characteristics, which provides some interesting relationships for modeling and useful information for parameter and state estimation.

Chapter 4

Background on system identification and CLM2.0 description

4.1 Introduction

Reproducing the observed behavior of a system by a computer coded model, requires a decent system identification. Through analysis of signals entering and leaving the system, relationships between physical variables in a (natural) system can be set up and mapped onto mathematical structures, like simple algebraic equations or more complex systems of partial differential equations (PDE). In the construction of a mathematical model for natural phenomena, a choice of the model structure and of the model's parameters should be made. Some principles of system modeling are explained by De Keyser [2001] and an example for a hydrological problem is given by van Loon and Keesman [2000]. System identification implies the use of a priori knowledge, for engineering systems based on laws of physics, to determine the structure of a system (white models). Additionally, measurements are used to obtain the necessary a posteriori knowledge of the system under test. Strictly black box models are built using only measurements without any hint about the structure, but are hardly feasible and unrealistic, and thus most of these models end up to be rather gray. The idea of system identification is summarized in figure 4.1. Once the model structure is known, identification involves the estimation of possibly the order, the state and the parameters of the system. During the process of system identification, information from observed data is essentially assimilated in order to build the model, and thus this process can be seen as a particular form of data assimilation sensu largo.

If the structure of a model is (assumed to be) known in advance, parametric identification methods can be used to fully describe a system. In this work it is assumed that with the model structure, the order of the system and the structure (not necessarily the values) of the state vector are known. Calibration of a model is needed in this research prior to filtering to assure a realistic transition function and to reduce the model error as close as possible to a zero mean white Gaussian noise. Once the model structure and parameters are determined, the quality of the resulting model should be judged through validation with a dataset that is independent of the calibration data set.

First some model structures are discussed and some literature on parameter estimation or calibration is reviewed. Next, the specific system model of the CLM2.0 and some changes introduced in this study are discussed.

4.2 Model structure

Soil moisture processes have been represented by a wide variety of models, ranging from simple 1D Darcy-based equations for profile simulations only to very complex land surface models. Research on global scale earth processes mostly involves the use of global General Circulation Models (GCM). Land surface processes are part of the total of global processes controlling the earth [Pitman, 2003] and represented in Land Surface Models (LSM). Most LSMs are Soil-Vegetation-Atmosphere Transfer (SVAT) models, where the vegetation is mostly not a dynamic component. Recently, coupling of hydrological or SVAT models with vegetation models has received some attention [Arora, 2002; Pauwels *et al.*, 2006]. In this study, the focus is on hydrological land surface processes within the thin layer ranging from the lower atmosphere (troposphere) to a few meters into the soil. Depending on their structure, GCMs and LSMs can be categorized, based on different features:

- Theoretical models are completely based on laws of physics (differential equations), controlled by equations of conservation of mass and energy. They are founded on rigorous numerical solutions of partial differential equations.

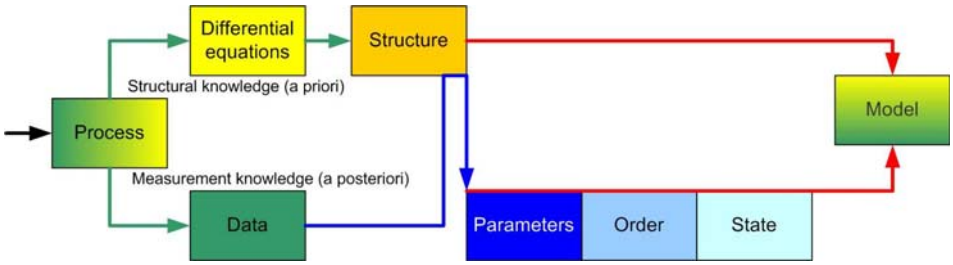


Figure 4.1: System identification.

Given known boundary conditions, the complete mathematical model is formed. These models are also called physical process models or physically based models and can be seen as white models. In contrast to these models, experimental or empirical models are built on the analysis results from input and output signals and are often black box models. They might be less complex, but at the risk of losing the internal structure of the processes. Most LSMs are a combination of both: most existing theoretical models simplify the physical system and often include obviously empirical components and therefore they are considered conceptual models. Conceptual models rely on some schematization of the watershed's dynamics (conceptualization), with parameters being related to physical properties in the catchment. SVAT models are typically categorized as physically based, even though some empirical relationships are included. Grayson *et al.* [1992b] gave a critical review of the use of physically based distributed models in hydrology.

- A lumped system is one in which the dependent variables of interest are only a function of time. In general, the model solves a set of ordinary differential equations (ODEs). A distributed system is one in which all dependent variables are functions of time and one or more spatial variables. In this case, the model will typically be solving partial differential equations (PDEs). Lumped models do not explicitly take into account the spatial variability of inputs, outputs, or parameters, while distributed models do. Obviously, lumped models are easier to calibrate than distributed ones.
- Other typical classification criteria for models in general are discrete-time versus continuous-time, linear versus non-linear, input-output versus state-space, stochastic versus deterministic, parametric versus non-parametric and dynamic versus static models. Most LSMs are dynamic, discrete-time, non-linear, state-space, deterministic and parametric.

It is recognized that different model structures can lead to similar results. It is even possible that bad or wrong representations of watershed processes lead to good model results [Grayson *et al.*, 1992a]. In that case, the problem is that internal estimates of variables will probably be erroneous. Hogue *et al.* [2006] evaluated different LSMs and found that additional complexity did not equal to improved performance. A comparison of different soil moisture modeling schemes was done by Shao and Henderson-Sellers [1996] in the Project for Intercomparison of Land surface Parameterization Schemes (PILPS) project. The PILPS project [Pitman and Henderson-Sellers, 1998] aimed to foster an evaluation of the next generation of land surface schemes and to coordinate the evaluation of land surface schemes in their different applications.

In view of the Distributed Modeling Intercomparison Project (DMIP) of the Hydrology Laboratory of the National Weather Service (NWS) of the National Oceanic and Atmospheric Administration (NOAA), which is encouraging the use of spatially distributed data to improve model results, Ajami *et al.* [2004] compared the performance of a lumped to a semi-distributed version of a model and concluded

that for semi-distributed models the improvement of the simulation capability for integrated runoff at the outlet was not significant yet, but that the use of such models should be preferred to provide information at interior points of the modeled catchment. They also stated that moving from lumped to distributed models creates more complexity in modeling and in calibration and consequently more uncertainty in results. The uncertainty is also included in the distributed input, representation of processes and the increased number of parameters.

Most LSMs used in GCMs view the soil column as the fundamental hydologic unit, ignoring the role of e.g. topography on spatially variable processes [Stieglitz *et al.*, 1997] to limit the complexity and computations for these coupled models. Some efforts [Shaman *et al.*, 2002; Yang and Niu, 2003; Niu and Yang, 2003] have been done to include e.g. topography in such LSMs for GCMs, by incorporation of the TOPMODEL concept [Beven and Kirkby, 1979].

During the last decades, LSMs were built with a higher degree of complexity in order to better represent land surface atmosphere interactions within GCMs or to meet the need for knowledge of the local state and processes in, for example, environmental or agricultural management studies. This includes e.g. the treatment of more physiological processes, the improvement of the representation of subgrid heterogeneity and the development of distributed models. Practically, these advanced LSMs can/have not been used yet operationally coupled in GCMs, but relatively simple distributed LSMs have been used for agricultural and environmental purposes, such as the monitoring of pollution or nutrients [Quinn, 2004; Heathwaite *et al.*, 2005].

Ideally, an improved process representation (system model structure) should result in parameters that are easier to measure or estimate. However, a more complex process representation results in more parameters to be estimated and several authors [Beven, 1989; Duan *et al.*, 1992; Yapo *et al.*, 1998; Franks *et al.*, 1999] stated that LSMs are over-parameterized given the data typically available for calibration. In practice, effective parameter values are best derived through calibration, instead of using measured values, as they will depend on the scale [Feddes *et al.*, 1993; Beven, 1995; Refsgaard, 1997; Vrugt *et al.*, 2004], the values of other parameters (multi-parameter interactions) and on the model structure. It was found that measured parameters differed considerably from 'effective' parameters [Mertens *et al.*, 2005]. One of the most obvious causes is that models simulate processes at grid cells with an area of a (some) squared meter(s) to several squared kilometers, while the determination of parameters happens mostly on a smaller resolution. Furthermore, Davis *et al.* [1999] illustrated that different measurement techniques yielded different parameter estimates and the resulting differences in parameter values had a large effect on model results. However, inclusion of measurement information on the parameters may be valuable, as illustrated by Mertens *et al.* [2004], who estimated effective parameters with either a wide parameter range or the joint probability distribution of measured parameter values as prior information. Inclusion of prior information resulted in more realistic parameter values.

The inclusion of the scaling problem (aggregation/disaggregation) [Beven, 1995] in the structure of earth system models is a topic of current research. While GCMs are typically developed and run at the large grid scale, LSM operate often at the small grid scale or point scale. Most LSM do not contain an adequate up- or down-scaling procedure, and consequently separate calibration and validation tests should be carried out every time the grid size is changed. The use of global parameters will result in error in predicting local responses at points with some particular characteristics. This also means that one should be critical in using parameter values reported in literature.

While the majority of calibration studies have been concerned with 'lumped' applications, more and more research on calibration of distributed models has been reported, mostly however after a drastic initial reduction of number of parameters. In most cases only a limited soil layering was dealt with and several parameters were kept constant in space. Efficient and effective fully distributed calibration is still a topic of research (which will also be addressed in this work) and the use of fully distributed parameterization may not always imply improved model results. Boyle *et al.* [2001] did not find an improvement in model results when using spatially distributed surface characteristics (model parameters). Houser *et al.* [2001] used spatially variable soil and vegetation information in a distributed model and found that the model produced unrealistic simulations. This was likely caused by discretely assigning a single parameter set to a given area and to shortage of information on the real spatial variability of the parameters. Eckhardt *et al.* [2005] added a module to a global optimization algorithm to enable the user to formulate constraints and interdependencies between the model parameters during the optimization, while preserving the spatial patterns of parameters. It should be remarked that a distributed model structure serves rather to meet requirements of the model use, e.g. the understanding of processes at a specific point, than to improve the overall modeling performance.

The calibration of black (gray) box models, simply relating input to output without any attempt to describe processes occurring in the watershed, has been quite well described and found to be efficient and successful (e.g. least squares method, instrumental variables method, prediction error method [De Keyser, 2001]). Examples of system structure identification and system parameter identification of multiple input-output time series models are given by Cooper and Wood [1982a, 1982b]. Furthermore, black-box models can be easily formulated in a state-space framework, allowing for straightforward applications of techniques like Kalman filtering [O'Connell and Clarke, 1981].

4.3 Model parameters

4.3.1 Selection of optimization method

As mentioned above, in the study at hand the model structure is predefined and parametric identification is the last step to identify the complete system. The objective of calibration, parameter estimation or parameter optimization of any model for a physical system is to identify the optimal values of some parameters in a model. This is done by feeding the model with input data, and comparing the computed output variables to observations in the physical system. The difference is to be minimized, with the model parameters as the independent control variables. Several evaluation criteria can be used to capture this discrepancy or the goodness-of-fit of the model (cfr. infra). Traditional hydrologic calibration has been performed manually using a trial-and-error parameter adjustment method [Boyle *et al.*, 2000]. However, this is time consuming, very subjective and not necessarily optimal. Therefore, today most studies take advantage of automatic calibration procedures, where the calibration problem is mostly seen as an optimization problem.

The optimization problem can be posed as finding the optimal estimate for the parameter vector, that minimizes a certain objective function. As the value of the parameter vector is limited by a finite interval of the Euclidean space, this is a constrained optimization problem. The parameter space is usually defined as a hypercube by specifying upper and lower limits, or as a hyperellipsoid, using prior knowledge about the correlation between different parameters [Kuczera, 1997]. A hyperspace is a multi-dimensional space beyond the three dimensions, and therefore it is not easily represented. However, the constrained problem can easily be transformed to an unconstrained optimization problem by introducing a penalty function, that takes very high values for parameter values outside the finite predefined interval. This is beneficial as most optimization algorithms are developed for unconstrained problems. Typically, these algorithms can be divided into local and global optimization algorithms. Local search involves finding a single extremum of an objective function. If an analytical expression for this function is available, local search methods (e.g. steepest descent, conjugate gradient, quasi-Newton method) are often gradient methods, using derivatives of the objective function or adjoint equations to facilitate the computation for more complicated models (e.g. White *et al.* [2003] for a physically based distributed model). Otherwise, direct search methods are used (e.g. downhill simplex, rotating directions, direction set method). These methods are often applied without checking for unimodality, leading to suboptimal results. It should be recognized that in most hydrological applications, the assumption of a single extremum in the response surface of an objective function does not hold (definitely not in conceptual models) [Duan *et al.*, 1992; Kuczera, 1997] and consequently, global search methods are needed to find the best optimum. Examples of such methods are set covering methods, random search methods, methods based on multiple local searches, population-

evolution based methods, such as genetic algorithms, shuffled complex evolution (SCE, Duan *et al.* [1992]) and particle swam optimization (PSO, Kennedy and Eberhart [2005]). A summarizing overview of several methods in a hydrological context and an example of automatic calibration of a groundwater model, using global optimization, was given by Solomatine *et al.* [1999]. More applications of global search methods in hydrological models were reported by Kuczera [1997], Sorooshian *et al.* [1993], Senarath *et al.* [2000], Eckhardt and Arnold [2001] and many others.

Remark that in essence, these optimization methods are exactly the same techniques as used for variational data assimilation for state estimation. For calibration, an objective function is defined with the parameter vector as a control vector which is to be optimized, and the best parameter vector is found by minimizing the objective function. For variational assimilation methods, the procedure is identical, except that the control vector is the state vector, instead of the parameter vector. If the control state vector would be extended with parameters, then a weak constraint variational assimilation procedure is implied.

In agrophysical hydrologic models, soil hydraulic parameters have often been derived through 'inverse modeling' of an observed variable [Hupet *et al.*, 2003; Vrugt *et al.*, 2004], as opposed to the forward modeling discussed until now. Through inverse modeling it is sought which parameters are best suited in a model to obtain a given set of observations, collected in a well defined experiment. For example, for groundwater models the water table depth is calculated as function of the conductivity of the soil. Through inverse modeling, an observed water table depth is put in the model and the corresponding conductivity will be calculated by the inverse model. A well known parameter estimation package, that is often used for groundwater modeling, is PEST (Parameter ESTimation, software developed by Watermark Numerical Computing). This package runs the model repeatedly and adjusts model parameters until the discrepancies between selected model outputs and a complementary set of measurements is reduced to a minimum in the weighted least-squares sense. Goegebeur and Pauwels [2006] reported that this procedure could fail for high observation errors, infrequent observations or erroneous initial parameter guesses.

4.3.2 Selection of parameters to be optimized

Before starting the calibration of a model, one should select those parameters for optimization that are difficult to measure or that are scale dependent. Further, it is well advised to perform some sensitivity analyses in order to select only those parameters for calibration that have a significant impact on the model results, to reduce the dimension of the parameter optimization problem. A priori knowledge, such as field measurements of some parameters or correlations between different parameters should be used to limit the amount and possible range of parameter values to calibrate on. Sensitivity analyses in general (e.g. for Biosphere-Atmosphere

Transfer Scheme (BATS) by Gao *et al.* [1996]) aim to investigate the responses of a model to perturbations in the parameter values, input variables and the (initial) state. The latter two will be discussed in section 4.4 and 4.5. The sensitivity to parameters can be tested by perturbing each parameter at a time, but this ignores the effect of multi-parameter interactions. Anderton *et al.* [2002] concluded from their sensitivity analysis with SHETRAN (extended version of the Système Hydrologique Européen (SHE), [Abbott *et al.*, 1986]) that interactions between parameters are complex and that changes in values of individual parameters must always be considered in the context of other parameters and the state of the system.

An example of a multi-objective sensitivity analysis approach is the Multi-Objective Generalized Sensitivity Analysis (MOGSA) [Bastidas *et al.*, 1999] to determine main influential parameters. Using this approach, Demarty *et al.* [2004] found that the sensitivity of a model to given parameters is highly dependent on climatic and environmental conditions. MOGSA was also used by Demarty *et al.* [2005] within a multi-objective calibration procedure in which the preferential ranges for the most influential parameters were determined, to allow contraction of the feasible parameter space.

Once the sensitive parameters are chosen, the range of possible values should be determined. Even for calibration runs with parameters chosen within a rather narrow range of observed or field values, a wide range of responses can be obtained [Anderton *et al.*, 2002]. Anderton *et al.* [2002] also suggested that incorrect estimation of model parameters not considered in the sensitivity study may cause parameter optima to be found in different parts of the parameter space.

4.3.3 Selection of objective functions

Different measurement series

LSMs simulate the evolution of different variables in the state vector and fluxes. All modeled processes are dependent on at least a few parameters. To find the optimal parameters for all those processes, calibration of these processes would ask for data on many variables involved. However, all information or measurements needed to find the real optimal parameter vector is never available causing that the problem is ill-posed. Simple solutions just consider a step-by-step calibration for the different variables, running the risk of a step-by-step compensation of errors.

Several authors reported improvements through multi-objective calibration by simultaneous use of observed data time series for different variables or at different sites. Franks *et al.* [1999] found that through conditioning a SVAT model on several heat flux objectives, a significant additional constraint of the feasible parameter space was achieved, resulting in a reduction of the predictive uncertainty. Gupta *et al.* [1999] reported that traditional single-criterion methods to estimate parameters in a land surface scheme were of limited value, while a multi-criteria approach was

effective in constraining parameter ranges into physically plausible ranges, when observations of at least one appropriate state variable and one flux were available. Anderton *et al.* [2002] demonstrated that the most appropriate parameterization of the SHETRAN subsurface flow model depended on the response (output variable) considered and on the criteria used to evaluate the model's goodness-of-fit to that response. McCabe *et al.* [2005] recognized the typical multi-output nature of models and found that combining different observation data streams for calibration helped to come up with a more robust process model and improved surface flux predictions.

Several authors [Refsgaard, 1997; Andersen *et al.*, 2001; Wooldridge *et al.*, 2001; Anderton *et al.*, 2002] made clear that the traditional calibration on outlet discharge only is not advisable, as evaluation of internal processes provided clear evidence of compensating errors (between modeled variables) within the model: validation at internal points showed significantly poorer results compared to the calibration sites. It can be concluded that attention is asked for multi-site and multi-variable calibration, which are specific kinds of multi-objective calibration.

In earlier work, Kuczera [1983b] warned for pooling of incompatible information. Different kinds of data should be consistent or in other words: parameters inferred from different kinds of data on the same system should not differ significantly. This refers to the need for data with minimal observation error.

Different measures of goodness-of-fit

The classical approach to calibration is to find the parameter vector that minimizes a measure of the difference between the observed and simulated data, i.e. the residuals. The residuals include both model error and error in the data. Measures of goodness-of-fit are objective functions that generate a response surface for a collection of parameter sets. Parameter estimation procedures are commonly based on the assumption of a correct model structure. This reduces the idea of calibrating a model in order to get rid of unknown/undefined model errors to the minimization of the parameterization error component in the model error. Typically, the expected probability of the data, given the model, is analyzed. The error is assumed to be observational [Romanowicz *et al.*, 1994, 1996]. However, in hydrology, mostly the observations are more correct than the model. Given the data, it is more interesting to know how likely different values of parameter are. This is given by a likelihood function. In the probability function, the parameter is assumed to be fixed, and the observations are assumed to be variable (and a function of the parameters). In the likelihood function, the parameter is a function of the data. In environmental modeling, there is usually more interest in the likelihood of a model as a simulator of the system, given the input and the observations, rather than the likelihood of the observations, given the model, inputs and the error model [Beven and Freer, 2001; Beven, 2000].

Several measures of goodness-of-fit have been proposed in literature, such as the

Root Mean Square Error ($RMSE$), Nash-Suttcliffe criterium (N_S), time series correlation coefficient (R) and (absolute) mean difference ($BIAS$), which all will be applied in later chapters. Several criteria were given by Beven [2000] and Gupta *et al.* [1998] listed some objective functions used by the National Weather Service of the U.S. for calibration of the Sacramento Soil Moisture Accounting (SAC-SMA) model. Alternative measures are often related to the variable under study, like hydrograph indices [Shamir *et al.*, 2005] to capture streamflow characteristics. Because of the Cramér-Rao [Cramér, 1946] bound, fundamental limits on parameter estimation performance can be derived for any parameter estimation problem where the parameter is not random.

Boyle *et al.* [2000] found that parameter sets resulting in a similar measure of goodness-of-fit caused very different simulations, pointing to the need for more criteria. Similarly, Madsen [2003] stated that a single objective measure is often inadequate to properly take into account the simulation of all characteristics a hydrologist is interested in, in order to evaluate the goodness-of-fit of the model. In a sensitivity study, Anderton *et al.* [2002] used for the discharge evaluation the N_S criterion and the ratio of the sum of the simulated and the observed discharge flow, while for the soil moisture reserve and the phreatic surface level, the correlation coefficient, the slope of the regression line and the difference between the means was used. This variety in measures allowed for a more complete analysis of the errors between simulated and observed values.

These indications that a single measure of goodness-of-fit does not suffice to properly calibrate models are linked with the fact that structural model error (e.g. from unrepresented scales) may sometimes be larger than observation error [Yapo *et al.*, 1996] and that the model error may not have the inherent probabilistic properties that can be exploited in the construction of an objective function. Therefore, Gupta *et al.* [1998] stated that there may not exist a statistically correct choice for the objective function and consequently no statistically optimal parameter set can be found. The 'multi-objective equivalence' of several parameter sets is commonly referred to as 'Pareto-optimal' or 'non-dominant'.

Gupta *et al.* [1998] suggested that the emergence of more powerful model calibration methods must include the recognition of the inherent multi-objective nature of the problem and explicitly recognize the role of model error. Therefore, they proposed to set the goal of calibration to minimizing the total residual to the measurement error instead of forcing the total residual to zero, which is the same as driving the model error to zero. They discussed that the classical option is to focus on the measurement errors and define an objective measure, taking into account the statistical distribution of these errors. Examples of this approach are the mean squared error (MSE) for homoscedastic independent error and maximum likelihood techniques for heteroscedastic independent error (HMLE) [Sorooshian and Dracup, 1980; Kuczera, 1983a, 1983b]. Homoscedastic errors are errors with similar variance for each magnitude of the variable (e.g. for each soil moisture value), while heteroscedastic errors have different variances. Another option is to ignore the statistical properties of the measurement errors. Then it may be

necessary to define multiple measures in order to drive every component of the residuals vector (focus on model error) to zero. In theory, this could ask for as many measures as there are components in the residual vector. In practice the number of measures will be significantly less. Consequently, multi-objective calibration is needed, even when only one kind (variable) of measurement data or output series is available.

A remedy to overcome potential model structural deficiencies is to parameterize this error [Yang and Michel, 2000], which introduces even more parameters into the calibration problem.

Combination of different objectives

The solution of a multi-objective calibration problem will not be a single unique set of parameters, but rather a Pareto-set of solutions (non-dominated solutions), as a result from various trade-offs among the objectives (different measured variables and measures of goodness-of-fit). Pareto procedures, such as Multi-Objective Complex Evolution (MOCOM, Gupta *et al.* [1998], Yapo *et al.* [1998]), Multi-Objective Shuffled Complex Evolution Metropolis (MOSCEM, Vrugt *et al.* [2003]) or Multi-Objective Particle Swarm Optimization (MOPSO, Gill *et al.* [2006]), search for the best parameters for several different objective functions and lead to a set of 'equally good' parameter sets: the Pareto solution space. The space increases with the number of objective functions. For practical applications a subjective choice has to be made on the single best set to use, but recently, Khu and Madsen [2005] proposed a preference ordering routine (POR) to reduce the number of possible solutions. Furthermore, the definition of several objective functions and the minimization/maximization require a lot of computational effort. Therefore, the reduction of several objective functions to a single global criterion is an interesting option to reduce the complexity of the mathematical minimization/maximization problem [Beven, 2000; Madsen, 2000a]. The objective functions are aggregated after transformation or scaling. This combination of objectives always involves some subjectivity. Some examples of multi-objective calibration studies are reviewed in next paragraph.

Houser [1996] performed a series of single-objective Monte Carlo (MC) calibrations of the TOPMODEL based Land Atmosphere Transfer Scheme (TOPLATS, [Famiglietti and Wood, 1994]) for several variables (soil moisture, soil heat flux, latent heat, sensible heat) at different sites and found that no common optimum set could be found for all objectives. As the size of Pareto-sets for the entire set of objective functions were excessive, the number of objective functions was reduced. Houser *et al.* [2001] delineated a 'noninferior' trade-off parameter region, by finding those MC points for which there did not exist at least one MC realization for which the predefined objective functions were better. From these 'noninferior' parameter sets, a best set was subjectively extracted. Madsen [2000a, 2000b] defined multiple objective functions to capture different aspects of the shape of hydro-

graphs and found that no single parameter set was able to optimize all objectives simultaneously and that a large variability was observed in the Pareto optimal parameter sets. He proposed a balanced aggregated objective function to determine a single best solution of the optimization problem, using a SCE algorithm. An Euclidean distance function was used in which all the objective functions were transformed such as to ensure that all were at the same distance from the origin, near the optimum. van Griensven and Bauwens [2003] considered water quantity and quality variables simultaneously, normalized the different objective functions according to cumulative probability distributions of the randomly sampled objective functions values in the parameter space and minimized the single global optimization criterion by the shuffled complex (SCE) algorithm. Madsen [2003] calibrated the distributed model MIKE SHE (extended version of SHE) using measurements of groundwater levels at several locations within a catchment and outlet runoff. After pooling for the groundwater levels over the different sites, 2 objectives were used, one for each variable. An aggregated measure was defined based on these 2 objectives through weighted averaging after an appropriate transformation. The SCE method was chosen as the optimization algorithm. Franks *et al.* [1999] combined normalized objective likelihood measures for each individual variable of heat flux through Bayes' equation in a generalized likelihood uncertainty estimation (GLUE) framework (cfr. infra). Schoups *et al.* [2005] applied for a subsurface water flow model a single-objective global optimization, in which three measurement types were combined in a single criterium and a three-objective global optimization, where the interdependence of the three measurement types was maintained. By the first method, they found parameters with small uncertainty bounds, while the second method resulted in a large parameter uncertainty in the Pareto set, indicating the possible inadequacies in the model structure.

4.3.4 Selection of calibration period

As discussed in the above sections, the selection of optimal parameters depends on the objectives used in the optimization problem and on the optimization technique itself. Parameters are also highly dependent on the calibration period. Franks and Beven [1997] discussed how even a simple SVAT model led to multiple acceptable parameterizations, when calibration data were limited to time scales of typical intensive field campaigns. Calibration data should contain information representing all the pertinent processes, i.e. longer calibration periods are likely to be beneficial to constrain the predictive uncertainty of a model. However, it is important to realize that the informativeness of the data is far more important than the amount used for calibration [Sorooshian *et al.*, 1983; Yapo *et al.*, 1996].

For automatic calibration of lumped models, operated at daily time scale, Yapo *et al.* [1996] and Gan *et al.* [1997] suggested a minimal calibration period of 8 years and at least one hydrological year of daily data, respectively. A hydrological year is a continuous 12-month period selected in such a way that overall changes in storage are minimal, so that carry-over is reduced to a minimum. These au-

thors also indicated that calibration for wet years is advised. Brath *et al.* [2004] noted that for lumped models the calibration period may be extended, because of the extremely short time needed for each simulation step, while for distributed models, the computation has to be carried out at each cell, which increases the simulation time and hence the time needed for calibration, because this process asks for multiple simulations. Since initial conditions did influence the values of the parameters (see section also 4.4), Brath *et al.* [2004] tested the influence of the length of the calibration period for a distributed model, starting from initial conditions obtained after some years of spin-up. They found that reducing the calibration period under 3 months deteriorated the model performances.

Often, calibration data do not display the full range of possible system dynamics and the information content may be restricted e.g. to particular surface conditions during a limited time period, which inevitably makes the resulting calibrated model less robust. A typical problem arises when the goal of modeling is to predict the effect of land use changes, where the determination of correct parameters for catchment conditions before and after land use change is needed [Wooldridge *et al.*, 2001].

In most studies the split sample calibration-validation has been applied, with the model being calibrated in off-line mode by assimilating the part of the available data and using the other data for validation. This becomes a clear limitation when hydrologic predictions must be generated from a previously ungauged watershed that has only recently been equipped with measurement instruments. Furthermore, this assumes a time-invariant system, which may not be a valid assumption.

Adaptive parameter estimation

Solving the calibration on-line, by recursively estimating the parameters, through assimilation of data as they become available, is highly likely to result in more optimal parameters and to reduce the parametric as well as the predictive uncertainty of models.

For gray box models, most parameter estimation methods also have a recursive version, allowing on-line parameter estimation. Cooper [1982] discussed the recursive least squares method, the instrumental-variable algorithm and the Kalman filter as recursive parameter estimation procedures and demonstrated an adaptive parameter estimation procedure, using an autoregressive moving average with exogenous input (ARMAX) model.

Brath and Rosso [1993] made clear that black (gray) box models are often preferred to conceptual ones, as both adaptive and non-adaptive calibration are more difficult for the latter, because of the often cumbersome structure and redundant parameterization. Nevertheless, they stated that conceptual models are better tools and investigated adaptive calibration of a conceptual rainfall-runoff model. Thiemann *et al.* [2001] presented a Bayesian recursive estimation approach for hy-

drologic prediction with a conceptual rainfall-runoff model, that could be used for simultaneous parameter estimation and prediction in operational setting. Katul *et al.* [1993] applied the Kalman filtering theory to determine the hydraulic conductivity function from a drainage experiment with a simplified Richard's equation. Rajaram and Georgakakos [1989] discussed recursive parameter estimation in conceptual hydrologic models by use of an augmented state extended Kalman filter. Franks and Beven [1997] proposed to incorporate more data, as they become available, into the likelihood estimates, using the GLUE methodology. Updating the likelihood distribution may be achieved through the application of Bayes' equation. Recently, the use of sequential techniques based on the Kalman filter have received increasing attention for parameter estimation in hydrology [Leng and Yeh, 2000; Boulet *et al.*, 2002; Roux and Dartus, 2005].

Yang and Michel [2000] stated that conceptual models are highly non-linear and cannot easily accommodate theoretically optimal methods such as e.g. Kalman filtering for updating parameters. They developed an alternative way to update parameters, starting from a reference parameter set representing the long-term behavior of the catchment and allowing only corrections that, after tracing backwards in time, produce a substantial effect at the current time and that are of such a nature that the model does not deviate too far from the main track.

The above filtering methods are basically traditional data assimilation techniques usually known in the context of state estimation, where data are assimilated when they become available, in order to update the model state. During calibration, the model is updated by updating the parameters and consequently improving the system representation. Through a combined state and parameter estimation the most promising results can be expected [Bras and Rodriguez-Iturbe, 1985; Hebson and Wood, 1985; Thiemann *et al.*, 2001; Vrugt *et al.*, 2005; Moradkhani *et al.*, 2005a, 2005b].

4.3.5 Selection of time and space resolution

It has been discussed already that parameters are very sensitive to the spatial scale and resolution on which they should be effective. Therefore, measured parameters may sometimes not be appropriate. Zehe and Blöschl [2004] remarked that the predictability of simulations is dependent on the scale, with increasing predictability for regional or global scales compared to field scales. Besides the spatial scale, also the temporal scale and resolution, or the calibration period and the magnitude of the time steps for simulations is of importance. Shamir *et al.* [2005] estimated parameters through analysis of streamflow characteristics at three time resolutions (monthly, yearly, record-extend). The posterior distributions of the model parameters derived at coarser time resolutions were used to sample model parameters for finer time resolutions.

4.3.6 Remaining uncertainty

Uncertainty

Even when a model is well calibrated, some uncertainty in model forecasts will always remain [Moore and Doherty, 2005]. It should also be clear that parameters take values, conditioned on the model structure (and error), initial conditions, input forcings, other parameters, the data error, the objective function used and the period of calibration. Anderton *et al.* [2002] reported that no single parameter combination could satisfactorily represent all aspects of a typical catchment hydrograph shape, possibly due to omissions or simplifications in the model process representation. Because of this uncertainty, Krzysztofowicz [2001] stressed that forecasts should always be stated in probabilistic, rather than in deterministic, terms. Best estimates of a state should always be accompanied with a quantification of uncertainty. One possibility is to generate an ensemble of realizations, as will be discussed in next chapter. The risk is that such ensembles tend not to cover the total uncertainty, but only 'known' uncertainties in some apparent uncertain variables or parameters, while a probabilistic forecast should quantify the total uncertainty.

The model will never be able to perfectly match the observed data, because of errors in both the model and the observations (cfr. supra). Model error apart from parameterization error may arise from simplifications, structural deficiencies, etc. Observation error is mostly caused by problems with sensors, such as calibration error [Hymer *et al.*, 2000], bad installation, etc. Both sources of error will also influence the state estimation in a later stage of data assimilation and will be discussed in extensu in later chapters.

The representation of model uncertainty led to the possibility for statistical analysis of model parameter uncertainty [Spear and Hornberger, 1980]. While much research has been done towards finding the most likely parameter set, somewhat less interest has been shown for the determination of the uncertainty of the found parameters [Kuczera, 1983a; Vrugt and Boutem, 2002]. Nevertheless, improved interpretation of parameter uncertainty would help to better explain the limits of our theoretical understanding of processes. In this respect the use of sequential data assimilation (e.g. Kalman filter) to estimate parameters is interesting, as it provides also an estimate of the uncertainty of the estimate. To the contrary, in variational data assimilation the calculation of the uncertainty is not directly included in the estimation, but it can be calculated separately.

Equifinality

Optimal system identification based on a limited amount of data does not necessarily result in a single best correct system description. The equifinality concept [Beven, 1993; Beven and Freer, 2001] holds the idea that there are many different

model structures and many different parameter sets for a chosen model structure that may be behavioral, equally probable or acceptable for reproducing the observed behavior of a system. This is mainly induced by over-parameterization of models. It also implies that incorrectly parameterized internal processes might still produce acceptable results in terms of discharge, as a result of compensating errors [Grayson *et al.*, 1992a]. Boyle *et al.* [2000] found that parameter sets that seemed to be equifinal, based on similar values for the measure of goodness-of-fit, were not similar in their ability to simulate different portions of studied hydrographs and could not be considered to be equifinal after further investigation. Duan *et al.* [1992] attributed part of the responsibility for the lack of unique optimal parameter solutions to optimization techniques that are not powerful enough to do the job.

Generalized likelihood uncertainty estimation (GLUE)

The GLUE [Beven and Binley, 1992; Beven and Freer, 2001] methodology for model identification rejects the idea that a single parameter set can be chosen to represent a physical system (equifinality). This methodology consists in a ensemble forecasting through MC simulations, using a sample of parameter sets from the behavioral model space, with each sample weighted according to its likelihood measure to estimate prediction quantiles. From these quantiles, uncertainty bounds can be calculated. The likelihood measure of the model performance is based on the modeling residuals. Several authors [Franks and Beven, 1997; Franks *et al.*, 1999; Romanowicz *et al.*, 1996] have experimented with the GLUE approach in hydrologic modeling.

4.4 Model state - initial conditions

The output of a deterministic model is completely determined if, besides the model structure and parameters, the initial and boundary conditions are known. Based on a sensitivity analysis to initial state conditions with the Biosphere-Atmosphere Transfer Scheme (BATS), Gao *et al.* [1996] warned that caution should be exercised when attempting to calibrate BATS using incomplete observation data or short term runs with erroneous values for the initial state. They found that errors in the initial state estimate can persist for a long time and that not all perturbations in initial model states would result in a model returning to dynamic equilibrium within a reasonable duration of time. Houser *et al.* [2001] found that spatially constant initial soil moisture initialization influenced the spatially distributed predictions and hence the calibration of an LSM. Senarath *et al.* [2000] stressed that calibration of physically based, two-dimensional distributed Hortonian hydrologic models were sensitive to the initial soil moisture field.

Analyses by Zehe and Blöschl [2004] indicated that the predictability of a model

is limited by the finite number of measurements to specify the initial soil moisture. Their simulations also revealed that the predictability was dependent on the average initial soil moisture state: an unstable range of 18 to 30 vol% resulted in relatively uncertain predictions, while for drier and wetter soil moisture, predictions were more certain. Further, different spatial patterns of initial soil moisture were reported to produce different results of runoff response, even if the spatial statistical distributions were the same [Merz and Bárdossy, 1998].

In practice, models have often been started from default, measured or guessed initial conditions until a stable situation was found and the initial state was balanced. This is called spin-up or warming-up of the model. The initial state can also be found through state estimation techniques as described in chapter 2. It should be noticed that through variational data assimilation for state estimation, the initial state is sought, for 4D-Var under the assumption of a perfect model and thus perfect knowledge of parameters. In calibration, the situation is usually just the other way around: the initial states are (assumed to be) known and the parameters are sought.

4.5 Model input forcings

Several authors studied the effects of spatial and temporal variability of mainly precipitation on model results and calibration. Spatially distributed precipitation data are currently widely available from weather radars. Boyle *et al.* [2001] found that the spatial representation of model input (precipitation) was, together with structural components of the model, the main cause of improvement of model performance for a distributed model. Ajami *et al.* [2004] also found some improvement when using a semi-distributed model using distributed precipitation. Similarly, Brath *et al.* [2004] found that model simulations were also satisfactory under the hypothesis of spatially uniform rainfall, provided that it was estimated based on sufficiently dense data in space, while a strong worsening was found when a spatially averaged value was estimated by rainfall data at a reduced density.

In most calibration studies, atmospheric forcing data were assumed to be accurate. Xia *et al.* [2005] found that forcing errors have few effects on the selection of optimal parameters when monthly evapotranspiration and runoff were calibrated, but that a significant effect was found when daily snow water equivalent was calibrated. Furthermore, the forcing errors did affect the uncertainty estimates of the parameters. Holman-Dodds and Bradley [1999] noticed that the temporal sampling may also affect calibration and that parameters compensate for the sampling effect, which detracts from the physical meaning of the parameters.

4.6 Community Land Model v.2.0 (CLM2.0)

In the study at hand the Community Land Model v.2.0 (CLM2.0) was used and therefore, its structure is presented here. The Community Land Model (CLM) is a global land surface model, developed through collaboration of different institutions and with the help of experts in different disciplines to provide the community with a free model that captures most of the best science currently available for land surface modeling. It is built on the fundamentals of ecological climatology. The initial CLM code was completed in late 1998 by combining the best features of three existing modular land models: the Biosphere-Atmosphere Transfer Scheme (BATS, Dickinson *et al.* [1993]), the Land Surface Model (LSM, Bonan [1996]), and the model developed at the Institute of Atmospheric Physics (IAP94, Dai and Zeng [1996]) in Beijing. Currently, several experts from different institutions are still contributing to the CLM project. Zeng [2003] reported on the experiences from this CLM project.

The CLM is the land model used in the Community Climate System Model (CCSM) and the Community Atmosphere Model (CAM) of the US National Center for Atmospheric Research (NCAR). CCSM is a global climate model with coupled sea ice, land, ocean and atmospheric components communicating through a coupler. CLM may be run in stand-alone mode, coupled to an atmospheric model, or fully coupled with CCSM, especially to investigate the effects of land surfaces on the climate and atmospheric chemistry.

CLM2.0, released in May 2002, is derived initially from the Common Land Model (CLM) and further updated versions became available during the course of this research work. At the time of writing, the most recent version is CLM3.0, released in June 2004 [Oleson *et al.*, 2004]. For this study we could not benefit from the more recent version, since the adaptations only result in minor changes in model output and the data structure has become less interesting for the applications in this study. In this work CLM2.0 is only used in off-line mode, without any coupling. The source code and global datasets to run CLM2.0 in off-line mode can be obtained from the WWW at http://www.cgd.ucar.edu/tss/clm/. The CLM2.0 User's Guide [Vertenstein *et al.*, 2002; Vertenstein and Hoffman, 2002] provides a manual on how to install and run the model. Dai *et al.* [2003] reported on extensive off-line tests with an initial version of CLM, using a variety of observational data.

For instructions on running CLM2.0 coupled to other CCSM2.0 components, one is referred to the CAM2.0 User's Guide and the CCSM2.0 User's Guide. Zeng *et al.* [2002] described the CLM and its impact on land surface climate after coupling to the NCAR Community Climate Model. Bonan *et al.* [2002] documented the effect of changes in the biogeophysics of CLM (resulting in CLM2.0) on the simulated climate through coupling with the NCAR CCM, and Kiehl and Gent [2004] discussed CCSM2.0 and a 1000 year control simulation. Radakovich *et al.* [2003] reported on the integration of CLM2.0 into the NASA/NCAR finite-volume Global Climate Model.

CLM2.0 was chosen for this research, because of the direct benefit from the expertise that collected the best science on earth processes and optimized the model structure and source code. The transparent modular structure of the code allowed a well organized implementation of the different algorithms discussed in this work. Its optimization for parallel processing allowed a lot of different experiments within a limited time. Furthermore, there is an active working group interested in applications of this relatively recent model. The main challenge for this research was the use of uncoupled land columns for a small land system, where in reality horizontal correlation was present. The objective was to deal with this shortcoming through data assimilation.

4.7 CLM structure

The overall structure of the CLM includes (1) the core single-point soil-snow-vegetation biophysical process code, (2) the land boundary data and (3) the scaling procedures required to interface atmospheric model grid-square inputs to land single-point processes. It is important and it will become apparent in the next sections that the core model can be tested with single point field data and that land surface datasets (derived from remotely sensed data and field data for respectively large and small scale applications) can easily be incorporated. Furthermore, the latest scaling procedures can be adopted. This will not be discussed in depth, as focus will be on the CLM in stand-alone mode, without coupling to an atmospheric or climate model. Because of spatial non-linearities of land processes (e.g. interaction with precipitation), calculation of grid-square average climate fields usually induces quantitative and qualitative inaccuracies [Beven, 1995]. Scaling procedures have an important function in many practical applications.

4.7.1 Horizontal structure

The CLM represents biogeophysical and other processes over a predefined grid by calculating water and heat fluxes and states for every grid cell separately, without any interaction between cells. Each grid cell can be subdivided into several tiles or patches, containing a single land cover type. These land cover types are each covering a fraction of the grid cell: vegetation, bare soil, wetland, lake, urban and glacier. The vegetated portion is further divided into patches of plant functional types (pft). This allows the definition of mixed life-form vegetation [Bonan *et al.*, 2002].

At every model time step, heat and water fluxes and state calculations are performed over each patch individually. There is no interaction between the patches, or in other words each patch maintains its own prognostic variables. By default, all patches within a grid cell have the same (grid cell) soil texture, soil color and corresponding physical properties. They respond to the same mean conditions

(forcings) of the overlying atmospheric grid cell, and the atmospheric grid cell responds to the areally-weighted fluxes from the underlying patches.

For some model applications described later, it was desirable that the patches differed in more characteristics than only in land cover. The possibility to change (perturb) the mean atmospheric forcings for every single patch was developed. It should be noted that for the small field scale application in this study, no information on the spatial distribution of atmospheric forcings is available. Consequently, all grid cells were overlaid with the same mean forcings. Spatial variation was induced through perturbation only. Furthermore, attributing different soil characteristics to each patch was introduced. Each patch can now be attributed different soil characteristics. This could be done by changing the texture for each patch. However, soil characteristics directly depend on experimental relationships with texture, which means that a certain soil type would always be characterized by the same characteristics (parameters), which is not necessarily true in the field. Therefore, variation of soil characteristics individually was introduced for each patch. Figure 4.2 gives a schematical overview of the structure, including the adaptations for this study. The modeling efforts were restricted to the 36 points (grid cells, profiles, land columns) for which observations were available. The maximum number of (pft-)patches per grid cell ever used in this study was 1500. Given the fact that the model runs for individual soil columns, the exact dimensions of the grid cells and patches was of no importance, because all variables and fluxes are calculated per unit volume or area.

4.7.2 Vertical structure

CLM has one vegetation layer, by default 10 vertical soil layers (changeable), and up to 5 snow layers (depending on the snow depth). For the soil layers, a CCM-like vertical discretization is used, with mesh points specified and interfaces located half way between 2 neighboring layers. For the study at hand, the choice for the depths of the different layer calculation nodes was mainly based on the measurement depths and the need for thin surface layers to assure numerical convergence. Experimental evidence showed that the choice of the discretizations had quite an effect on the model results (data not shown). In general, the numerical solution of the Richards equation converges as the resolution increases [Downer and Ogden, 2004]. Anderson and Woessner [1992] indicated that as a rule of the thumb, the thickness of a soil layer should not exceed 1.5 times the thickness the overlaying layer for numerical reasons. At the other side, a higher resolution and a more accurate solution results in higher computational costs. Therefore in most distributed models the number of layers is limited (e.g. 2 [Famiglietti and Wood, 1994] or 4 [Crow and Wood, 2003] in TOPLATS). Surface layers control rainfall partitioning and consequently the surface layer thickness and the upper boundary condition are of major importance. Fine discretizations at the top of the soil column were therefore advised and experimentally found to give more accurate results [van Dam and Feddes, 2000]. Coarser discretizations may be used at greater

depths. Downer and Ogden [2004] studied the sensitivity of model output to the estimates of model parameters, as related to the vertical cell size for the solution of the Richards equation. Values of parameters fitted for arbitrary discretizations run the risk of having no physical meaning. During calibration, compensation for errors in the numerical formulation (vertical discretization) may result in unrealistic parameters. Martinez *et al.* [2001] found that the water budget was very sensitive to the number of layers in the soil profile and that an insufficient number of layers may yield worse results than the use of an incorrect value of the saturated hydraulic conductivity. Another factor in favor of fine discretization is that the model becomes less sensitive to the initial soil moisture conditions and that the Richards equation formulation is able to correct errors in initial conditions properly, limiting the need for long time spin-up [Downer and Ogden, 2004].

Model soil depths were chosen at 2.5, 5, 10, 20, 30, 50, 80, 120, 150 and 180 cm depth. The corresponding layer thickness and the depths of the interfaces are given in the left plot of figure 4.2. It is clear that the layer thickness is unequal and that the difference in influence radius for the soil moisture values will be a source of representativeness error in the comparison of model results with observations.

4.7.3 Time-integration scheme

The model is integrated forward with a constant hourly time step. Calculated soil moisture fluxes may depend on the time steps used [van Dam and Feddes, 2000], just as noted for the space steps in the above section. After the model initialization, a time loop starts for the calculation of surface fluxes and the update of state variables for all patches. In the time loop first atmospheric forcings are processed and then the solution is split into an energy balance and a water balance phase in a very modular structure. For an overview on the code structure, the reader is referred to Vertenstein *et al.* [2002], figure 4.3 and the remainder of this chapter.

4.8 CLM physical processes

The description of the physical processes in this section is based on a thorough analysis of the source code, the technical documentation and user's guide of Dai *et al.* [2001], and the manual of CLM3.0 [Oleson *et al.*, 2004]. This discussion aims at clarifying where all parameters, which will be calibrated or perturbed in next chapters, occur within the model structure. Tunable parameters and constants are typed in bold fonts. Some constants in the default code were turned into tunable parameters for this study, because the modeled soil moisture was found to be sensitive to these parameters. These 'new' parameters are indicated with a ♠ symbol. Together with the definition of the different variables, the names used in the source code are given in small fonts, and an indication of the main

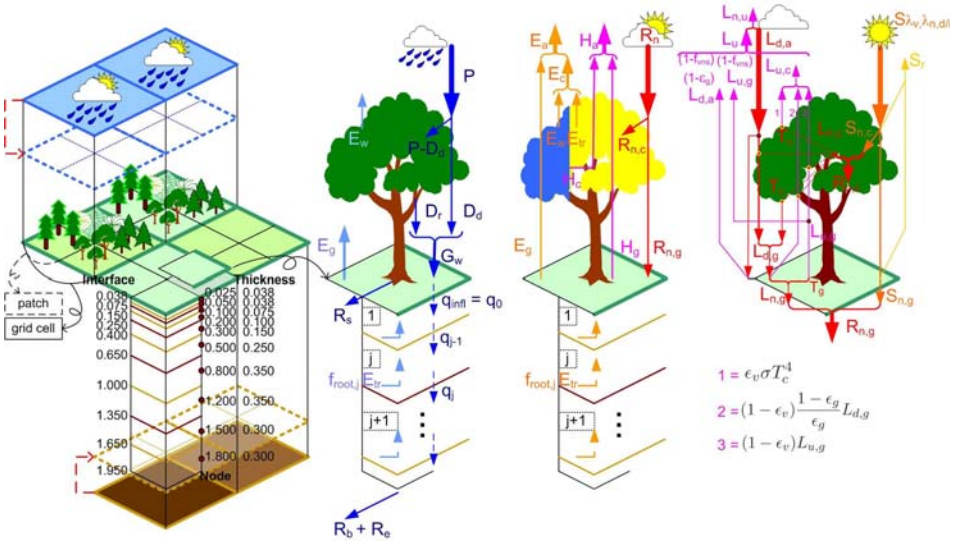


Figure 4.2: Far left: CLM2.0 structure with indication of grid cells and patches. The dashed lines represent the changes to the original structure: soil parameterization and forcings were assigned to patches instead of grid cells. Middle left: water balance in CLM2.0. Middle right: energy balance with indication of a wet (blue) and dry (yellow) vegetation part. Far right: radiation components. The symbols are explained in the text in section 4.8.

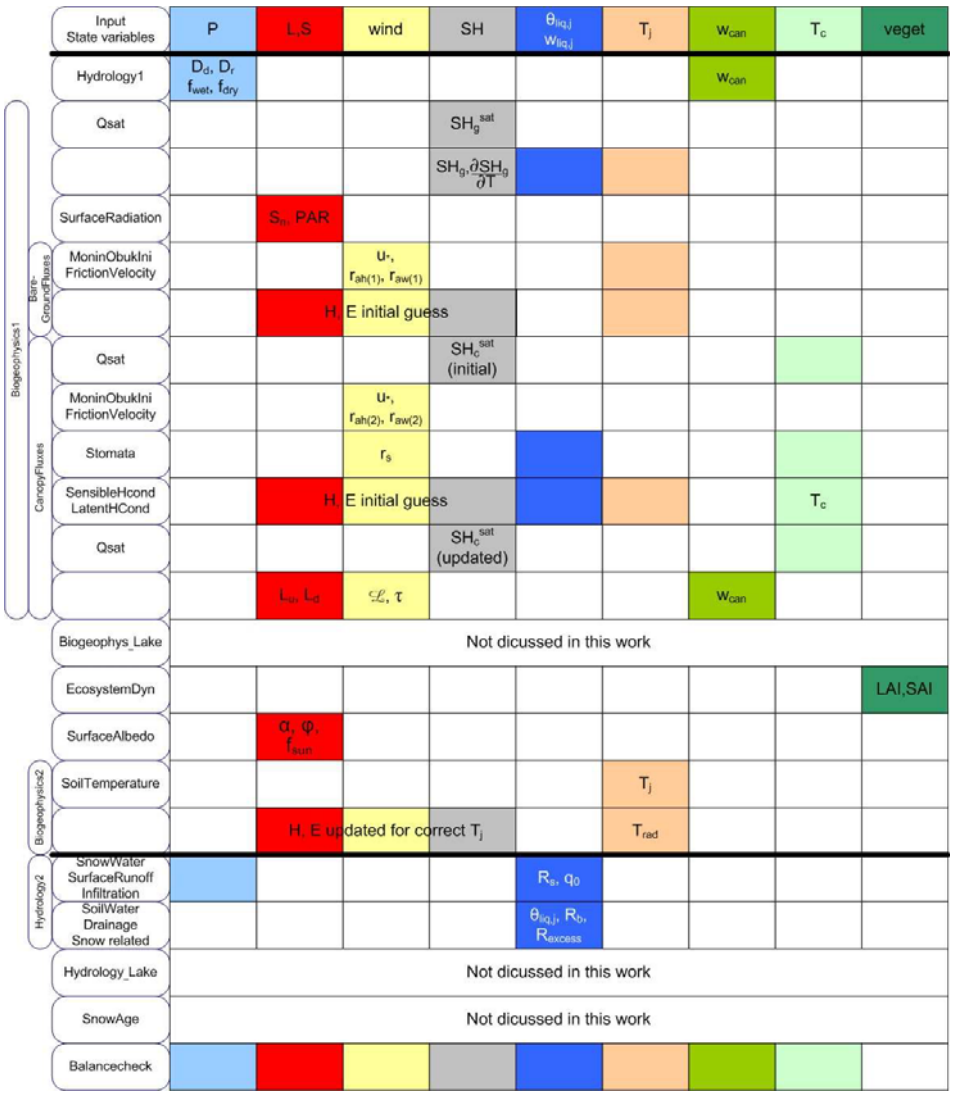


Figure 4.3: CLM2.0 modules (left) called in the driver module. Determination of variables is indicated by the name of the variable in a colored box, direct use of a variable is indicated by an empty colored box. All symbols are explained in the text.

module where these variables are used. It was assumed that each vegetation tile was completely covered by vegetation. Processes related to frozen water and phase change are not discussed in detail. All symbols are summarized at the end of this chapter, and are only for CLM2.0 variables in particular.

4.8.1 State variables

Each land surface patch has 6 prognostic variables, associated with the energy balance equations and the water balance equations: (1) canopy temperature (T_c), (2) temperature at each node of soil/snow (T_j), (3) canopy interception water store (w_{can}), (4) mass of water within each layer of soil/snow ($w_{liq,j}$), (5) mass of ice within each layer of soil/snow ($w_{ice,j}$) and the (6) snow layer thickness (Δz_j).

4.8.2 Water fluxes and states

The water balance is defined by mass conservation equations and calculated for each patch individually. It is driven by precipitation input and by evapotranspiration through radiation and vegetational processes.

Canopy water (Hydrology1.F90, CanopyFluxes.F90)

The water balance over a canopy control volume yields for the canopy interception water store, w_{can} [mm] (h2ocan, Hydrology1.F90, CanopyFluxes.F90) of a patch that is not buried with snow:

$$\boxed{\frac{\partial w_{can}}{\partial t} = P - D_d - D_r - E_w \ \ [\mathrm{kg.m^{-2}.s^{-1}}]} \tag{4.1}$$

with

P	=	precipitation [mm.s^{-1}]
D_d	=	direct throughfall [mm.s^{-1}]
D_r	=	drainage of water from foliage and stem [mm.s^{-1}]
E_w	=	evaporation from wet foliage [mm.s^{-1}]

- P: the precipitation contains the total of all liquid rain and solid snow falling out of clouds. (prcp = forc_rain + forc_snow, Hydrology1.F90)
- D_d: the direct throughfall of precipitation through the gaps in vegetation is given by (qflx_through, Hydrology1.F90):

$$D_d = P \cdot \exp[-0.5(LAI + SAI)] \ \ [\mathrm{mm.s^{-1}}] \tag{4.2}$$

  and the intercepted precipitation is given by $P - D_d$ (qflx_prec_intr, Hydrology1.F90), with

LAI = one-sided leaf area index with burying by snow [-] (elai,EcosystemDyn.F90)
SAI = one-sided stem area index with burying by snow [-] (esai,EcosystemDyn.F90)

The LAI and SAI are interpolated based (EcosystemDyn.F90) on the given monthly LAI, ***MONTHLY_LAI***, and SAI, ***MONTHLY_SAI*** (mlai, msai), the canopy top height [m], ***MONTHLY_HEIGHT_TOP*** (htop), and the canopy bottom height [m], ***MONTHLY_HEIGHT_BOT*** (hbot).

- D_r: the drainage or outflow of water stored on foliage and stems is given by (qflx_candrip, xrun, Hydrology1.F90):

$$D_r = \frac{w_{can} - w_{cmax}}{\Delta t} \ [\mathrm{mm.s^{-1}}], \text{ if } w_{can} > w_{cmax} \quad (4.3)$$

$$D_r = 0 \ [\mathrm{mm.s^{-1}}], \text{ if } w_{can} \leq w_{cmax} \quad (4.4)$$

with
w_{cmax} = maximum water capacity of the canopy [mm]
= $\boldsymbol{dewmx} \cdot (LAI + SAI)$ (h2ocanmx, Hydrology1.F90)
$\boldsymbol{dewmx}$ = maximum dew [mm], parameter $\boldsymbol{dewmx}$.
Consequently, the resulting $w_{can} = w_{cmax}$, if $D_r > 0$.

- E_w: the evaporation from the wetted fraction of the foliage (qflx_evap_veg-qflx_tran_veg, CanopyFluxes.F90) is determined by the air specific humidity and will be discussed in section 4.8.3 on the heat balance. It is through the evaporation that the heat and water balances of a land system are coupled.

Based on w_{can}, the wet and dry fractions of the canopy, f_{wet} and f_{dry}, (fwet, fdry, Hydrology1.F90), which are required for the surface albedo, are calculated by:

$$f_{wet} = \left(\frac{w_{can}}{w_{cmax}}\right)^{\frac{2}{3}} \quad (4.5)$$

$$f_{dry} = (1 - f_{wet})\frac{LAI}{LAI + SAI} \quad (4.6)$$

Soil water (Hydrology2.F90)

Soil moisture can be expressed in several ways. The liquid/ice mass in layer j is given by $w_{liq,j}$ and $w_{ice,j}$ [kg .m^{-2}=mm] (h2osoi_liq(j), h2osoi_ice(j)). The liquid/ice mass is related to volumetric soil moisture/ice, $\theta_{liq,j}$ and $\theta_{ice,j}$ (vol_liq(j), vol_ice(j), h2osoi_vol(j)), by:

$$\theta_{liq,j} = \frac{w_{liq,j}}{\Delta z_j \cdot \boldsymbol{\rho}_{liq}} \ [\mathrm{m^3.m^{-3}}] \quad (4.7)$$

$$\theta_{ice,j} = \frac{w_{ice,j}}{\Delta z_j \cdot \boldsymbol{\rho}_{ice}} \ [\mathrm{m^3.m^{-3}}] \quad (4.8)$$

with
$\boldsymbol{\rho}_k$ = density of constituent k [kg.m^{-3}], with $k = liq$ (liquid water) or ice
Δz_j = thickness of soil layer j [m], with the node index $j = 1, \ldots, N$ (nlevsoi)
= $z_{h,j} - z_{h,j-1}$, with $z_{h,j}$ the interface depths

The variable $\theta_{liq,j}$ (vol_liq, h2osoi_vol) is the state variable of particular interest for this study and will simply be noted as θ_j, if no confusion is possible.

The soil wetness (including ice) is defined as:

$$s_j = \frac{\theta_{liq,j} + \theta_{ice,j}}{\boldsymbol{\theta_{sat_j}}} \quad [-] \tag{4.9}$$

with
$\boldsymbol{\theta_{sat_j}}$ = volumetric soil water at saturation in layer j [m^3.m^{-3}],
parameter $\boldsymbol{watsat_j}$

Water flow in the vadose zone is predominantly vertical, and can be simulated as one-dimensional flow in many applications. In CLM2.0 the vertical water flow is calculated through a combination of the continuity equation for water (water mass balance) and Darcy's law. This approach approximates a 1D Richards equation. The vertical water transport is governed by infiltration, surface and subsurface runoff, gradient diffusion, gravity and root extraction through canopy transpiration.

Integration of the liquid water mass balance over a soil control volume (soil layer j of thickness Δz_j) yields:

$$\boxed{\frac{\partial w_{liq,j}}{\partial t} = (q_{j-1} - q_j) - f_{root,j} E_{tr} + (M_{ice,liq} \Delta z)_j \quad [\text{kg.m}^{-2}.\text{s}^{-1}]} \tag{4.10}$$

with
q_j = water flow between neighboring layers,
positive downward [kg.m^{-2}.s^{-1}]
$f_{root,j} E_{tr}$ = water extracted by plant roots, transpiration [kg.m^{-2}.s^{-1}]
$M_{ice,liq}$ = rate of phase change (ice to liquid) [kg.m^{-3}.s^{-1}]

- q_0: at the surface, the flow for bare soil is given by the infiltration (qflx_infl, Infiltration.F90):

$$q_0 = G_w - R_s - E_{g,Surf} \quad [\text{kg.m}^{-2}.\text{s}^{-1}] \tag{4.11}$$

with
G_w = sum of effective precipitation and snowmelt [kg .m^{-2}.s^{-1}]
R_s = surface runoff [kg .m^{-2}.s^{-1}]
$E_{g,Surf}$ = soil surface evaporation [kg .m^{-2}.s^{-1}]

  - G_w: the total amount of liquid water that really reaches the ground (qflx_top_soil, SnowWater.F90) is composed by throughfall, drip from canopy (qflx_rain_grnd, Hydrology1.F90) and snowmelt (qflx_snomelt, SnowWater.F90).

$$G_w = D_d + D_r + S_m \quad [\text{kg.m}^{-2}.\text{s}^{-1}] \tag{4.12}$$

with
S_m = snowmelt [kg .m^{-2}.s^{-1}]
For snow covered soil, the infiltration is given by $q_0 = S_m - R_s$.

– R_s: the total surface runoff (qflx_surf, SurfaceRunoff.F90) is the sum of surface runoff from saturated and unsaturated (qflx_drain_wet, qflx_drain_dry) regions:

$$R_s = (1 - f_{sat})\bar{w_s}^4 G_w + f_{sat} G_w \quad [\text{kg.m}^{-2}.\text{s}^{-1}] \tag{4.13}$$

with
f_{sat} = partial contributing area, saturated fraction [-] (fcov)
= $\boldsymbol{w_{fact}} \cdot \exp(-z_w)$
$\boldsymbol{w_{fact}}$ = determined by the distribution of the topographic index, parameter $\boldsymbol{wtfact}$ [-] (wtfact)
z_w = mean water table depth (zwt)
= $\boldsymbol{f_z}(z_{h,N} - \sum_{j=1,N} s_j \Delta z_j)$ [-]
$z_{h,N}$ = bottom depth of the lowest soil layer [m] (zi(nlevsoi))
$\boldsymbol{f_z}$ = water table depth scale parameter [m^{-1}], parameter $\boldsymbol{fz}$ (fz)
s_j = soil wetness (including ice) [-]
$\bar{w}_s$ = soil layer thickness weighted soil wetness [-] in the top $\boldsymbol{NwRs}$ layers
= $\frac{\sum_{j=1,\boldsymbol{NwRs}} s_j \Delta z_j}{\sum_{j=1,\boldsymbol{NwRs}} \Delta z_j}$ [-] (wmean)
$\boldsymbol{NwRs}$ = last top layer contributing to the calculation of surface runoff [-], parameter $\boldsymbol{NwRs}$
Note that the top soil layer is impermeable, if the effective porosity ($\boldsymbol{\theta_{sat_j}} - \theta_{ice,j}$) is less than a predefined parameter $\boldsymbol{wimp}$ [-]. In that case, $\bar{w}_s$ reduces to 1 and all water reaching the soil surface runs off.

– $E_{g,Surf}$: the evaporation from bare soil (qflx_evap_grnd, Biogeophysics2.F90) is determined by the profile of the humidity of the air and by the amount of water in the upper soil/snow layer (see section 4.8.3).

- q_N: at the bottom of the soil column, the flow is simply given by the hydraulic conductivity, i.e. assuming no change of soil matrix potential with depth, as for free drainage.

- q_j : the water flow from one soil layer to another is determined by Darcy's law:

$$q_j = -K_j \frac{(\psi_{j+1} - \psi_j) - (z_{j+1} - z_j)}{z_{j+1} - z_j} \quad [\text{kg.m}^{-2}.\text{s}^{-1}] \tag{4.14}$$

with
z_j = node depth of layer j, positive, increasing downward [mm]
$\psi_j - z_j$ = hydraulic height [mm], with $-z_j$ the gravitational potential, with the reference elevation at the soil surface
K_j = unsaturated hydraulic conductivity in layer j [mm.s^{-1}]
ψ_j = matrix potential of soil at layer j [mm] (< 0 in the unsaturated zone)

- $f_{root,j}E_{tr}$: the water extracted by plant roots is removed from the soil by transpiration (qflx_tran_veg). For each layer the effective root fraction $f_{root,j}$ (rootr(j)) is based on plant physiological characteristics (root distribution) and the soil matrix potential:

$$f_{root,j} = \frac{f_{r,j} w_{LT}(j)}{10^{-10} + \sum_{j=1,N} f_{r,j} w_{LT}(j)} \tag{4.15}$$

$$w_{LT}(j) = \frac{\boldsymbol{\psi_{max}} - \psi_j}{\boldsymbol{\psi_{max}} + \boldsymbol{\psi_{sat}}} \tag{4.16}$$

with
- $f_{r,j}$ = root fraction within soil layer j [-] (rootfr), calculated based on a pft-parameter for root distribution ***roota_par***
- $w_{LT}(j)$ = transpiration restricting factor [-] (rresis, CanopyFluxes.F90, SoilWater.F90), ranging from 0 at the permanent wilting point to 1 at saturation
- $\boldsymbol{\psi_{max}}$ = maximum value of negative of leaves potential before dessication or wilting point potential [mm], parameter ***smpmax*** (negative value)
- $\boldsymbol{\psi_{sat}}$ = saturated soil water potential [mm], parameter ***sucsat*** (positive value)
- ψ_j = matrix potential limited by $\boldsymbol{\psi_{max}}$ [mm]

The denominator of Eq. (4.15) is defined as the soil water transpiration factor (0 to 1) [-] (btran). The calculation of E_{tr} (qflx_tran_veg, CanopyFluxes.F90) is discussed in section 4.8.3. Following Dickinson *et al.* [1993] the transpiration cannot exceed the maximum transpiration that the vegetation can sustain. In CLM2.0 it is proposed that the capacity (btran) should be larger than 10^{-10} in order to have transpiration. Probably the parameter ***trsmx0*** [mm.s^{-1}] was introduced to refine this, but it is currently not used.

The numerical scheme (SoilWater.F90) is obtained by rewriting Eq. (4.10) for liquid water, which yields an expression for the soil moisture in soil layer j (of thickness Δz_j, not to be confused with the possible notation for spatial discretization) at time step $n+1$:

$$\frac{\Delta\theta_j^{n+1}\Delta z_j}{\Delta t} = [q_{j-1}^{n+1} - q_j^{n+1}] - f_{root,j}E_{tr} \quad [\text{mm.s}^{-1}] \tag{4.17}$$

with
- $\Delta\theta_j^{n+1}$ = $(\theta_j^{n+1} - \theta_j^n)$ [-], change in volumetric soil liquid water in layer j during time interval Δt (dwat)
- Δz_j = soil layer thickness expressed in mm, since $\boldsymbol{\rho_{liq},_j}$ is not included

which can be further numerically expanded as:

$$\begin{aligned}\frac{\Delta\theta_j^{n+1}\Delta z_j}{\Delta t} &= q_{j-1}^n + \frac{\partial q_{j-1}}{\partial\theta_{j-1}}\Delta\theta_{j-1}^{n+1} + \frac{\partial q_{j-1}}{\partial\theta_j}\Delta\theta_j^{n+1} \\ &\quad -q_j^n - \frac{\partial q_j}{\partial\theta_j}\Delta\theta_j^{n+1} - \frac{\partial q_j}{\partial\theta_{j+1}}\Delta\theta_{j+1}^{n+1} - f_{root,j}E_{tr} \end{aligned} \tag{4.18}$$

In general Eq. (4.18) can be written as:

$$a_j\Delta\theta_{j-1}^{n+1} + b_j\Delta\theta_j^{n+1} + c_j\Delta\theta_{j+1}^{n+1} = r_j \tag{4.19}$$

with

$$\begin{aligned} a_j &= -\left[\frac{\partial q_{j-1}}{\partial\theta_{j-1}}\right] \\ b_j &= \left[\frac{\Delta z_j}{\Delta t} - \frac{\partial q_{j-1}}{\partial\theta_j} + \frac{\partial q_j}{\partial\theta_j}\right] \\ c_j &= \left[\frac{\partial q_j}{\partial\theta_{j+1}}\right] \\ r_j &= [q_{j-1}^n - q_j^n] - f_{root,j}E_{tr} \end{aligned}$$

and

$$\begin{aligned} a_j &= 0, \text{ if } j = 1 \\ c_j &= 0, \text{ if } j = N \end{aligned}$$

with $a_j, b_j, c_j, r_j =$ amx(j), bmx(j), cmx(j), rmx(j), respectively, $q_{j-1} =$ qin, $q_j =$ qout, $\frac{\partial q_{j-1}}{\partial\theta_{j-1}} =$ dqidw0, $\frac{\partial q_{j-1}}{\partial\theta_j} =$ dqidw1, $\frac{\partial q_j}{\partial\theta_j} =$ dq0dw1, and $\frac{\partial q_j}{\partial\theta_{j+1}} =$ dq0dw2 (SoilWater.F90) .

Application of this equation to each node results in a tridiagonal matrix (Tridiagonal.F90) gives:

$$\begin{bmatrix} b_1 & c_1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & a_3 & b_3 & c_3 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & a_N & b_N \end{bmatrix} \begin{bmatrix} \Delta\theta_1^{n+1} \\ \Delta\theta_2^{n+1} \\ \Delta\theta_3^{n+1} \\ \vdots \\ \Delta\theta_{N-1}^{n+1} \\ \Delta\theta_N^{n+1} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_{N-1} \\ r_N \end{bmatrix} \tag{4.20}$$

Once the terms $\Delta\theta_j^{n+1}$ are known, $w_{liq,j}^n$ can be updated by:

$$w_{liq,j}^{n+1} = w_{liq,j}^n + \Delta\theta_j^{n+1}\Delta z_j \tag{4.21}$$

The terms in the coefficients, $\frac{\partial q_j}{\partial \theta_j}$ and $\frac{\partial q_j}{\partial \theta_{j+1}}$ are found by derivation of the flow q through the interfaces at the surface (Eq. (4.22)), the interior interfaces (Eq. (4.23)) and through the bottom interface (Eq. (4.25)):

$$q_0^{n+1} = q_0^n + \frac{\partial q_0}{\partial \theta_1}\Delta\theta_1 = \text{ infiltration} \tag{4.22}$$

$$q_j^{n+1} = q_j^n + \frac{\partial q_j}{\partial \theta_j}\Delta\theta_j + \frac{\partial q_j}{\partial \theta_{j+1}}\Delta\theta_{j+1} \tag{4.23}$$

$$\text{with } q_j = -K_j \frac{(\psi_{j+1} - \psi_j) - (z_{j+1} - z_j)}{z_{j+1} - z_j} \tag{4.24}$$

$$q_N^{n+1} = q_N^n + \frac{\partial q_N}{\partial \theta_N}\Delta\theta_N \tag{4.25}$$

$$\text{with } q_N = -K_N(0-1) = K_N \tag{4.26}$$

The respective derivatives are given by:

$$\frac{\partial q_0}{\partial \theta_1} = -\frac{\partial E_{g,Surf}}{\partial \theta_1} = S_d \Delta z_1 \tag{4.27}$$

$$\frac{\partial q_j}{\partial \theta_j} = -\left[\frac{K_j}{z_{j+1} - z_j}\frac{\partial(-\psi_j)}{\partial \theta_j} + \frac{\partial K_j}{\partial \theta_j}\frac{(\psi_{j+1} - \psi_j) - (z_{j+1} - z_j)}{z_{j+1} - z_j}\right] \tag{4.28}$$

$$\frac{\partial q_j}{\partial \theta_{j+1}} = -\left[\frac{K_j}{z_{j+1} - z_j}\frac{\partial(\psi_{j+1})}{\partial \theta_{j+1}} + \frac{\partial K_j}{\partial \theta_{j+1}}\frac{(\psi_{j+1} - \psi_j) - (z_{j+1} - z_j)}{z_{j+1} - z_j}\right] \tag{4.29}$$

$$\frac{\partial q_N}{\partial \theta_N} = \frac{\partial K_N}{\partial \theta_N} \tag{4.30}$$

with

$E_{g,Surf}$	=	ground evaporation [mm.s^{-1}]
S_d	=	extrapolates soil water dependence of ground evaporation, currently not used in the code and set to 0

- ψ_j: the soil matrix potential [mm] (smp(j), SoilWater.F90) and its partial derivative (dsmpdw(j), SoilWater.F90) at node depth z_j are given by:

$$\psi_j = \max\left[-\boldsymbol{\psi_{sat}}\left[\frac{\theta_j}{\boldsymbol{\theta_{sat_j}}}\right]^{-\boldsymbol{B}}, \boldsymbol{\psi_{min}}\right] \text{ [mm]} \tag{4.31}$$

$$\frac{\partial \psi_j}{\partial \theta_j} = -\psi_j\left[\frac{\boldsymbol{B}}{\theta_j}\right] \text{ [mm]} \tag{4.32}$$

with

$\boldsymbol{\psi_{sat}}$	=	saturated soil water potential [mm], parameter ***sucsat*** (positive value)
$\boldsymbol{\psi_{min}}$	=	restriction for minimal soil water potential [mm], parameter ***smpmin*** (negative value)
$\boldsymbol{B}$	=	parameter ***bws*** [-], defined by Clapp and Hornberger [1978]

- K_j: the unsaturated hydraulic conductivity (hk(j), SoilWater.F90) and its derivative (dhkdw(j), SoilWater.F90) at layer j are calculated based on the saturated hydraulic conductivity, $\boldsymbol{K_{sat_j}}$ (hksat(j), SoilWater.F90), which is assumed to decrease exponentially with depth (cfr. TOPMODEL concept) and calculated at the interface depth, $z_{h,j}$ (see figure 4.4):

$$\boldsymbol{K_{sat_j}} = \boldsymbol{K_{sat_0}} \cdot \exp(-\frac{z_{h,j}}{\boldsymbol{z_*}}) \quad [\text{mm.s}^{-1}] \tag{4.33}$$

with
$\boldsymbol{K_{sat_0}}$ = saturated hydraulic conductivity at the surface, parameter $\boldsymbol{xksat_0}$ [mm.s^{-1}]
$\boldsymbol{z_*}$ = length scale for K_{sat_j} decrease, parameter $\boldsymbol{hkdepth}$ [–]

The unsaturated hydraulic conductivity at layer j is then given by:

$$K_j = \boldsymbol{K_{sat_j}} \cdot s_j^{(2\boldsymbol{B}+3)} \quad [\text{mm.s}^{-1}] \tag{4.34}$$

Consequently, the hydraulic conductivity (hk(j)) and its partial derivative (dhkdw(j)) at interface depth $z_{h,j}$ are given by:

$$K_j = \boldsymbol{K_{sat_j}} \left[\frac{0.5 \cdot (\theta_j + \theta_{j+1})}{0.5 \cdot (\boldsymbol{\theta_{sat_j}} + \boldsymbol{\theta_{sat_{j+1}}})}\right]^{2\boldsymbol{B}+3} \quad \text{if } 1 \leq j \leq N-1 \tag{4.35}$$

$$= \boldsymbol{K_{sat_j}} \left[\frac{\theta_j}{\boldsymbol{\theta_{sat_j}}}\right]^{2\boldsymbol{B}+3} \quad \text{if } j = N \tag{4.36}$$

$$\tag{4.37}$$

$$\frac{\partial K_j}{\partial \theta_j} = \frac{\partial K_j}{\partial \theta_{j+1}} \tag{4.38}$$

$$= (2\boldsymbol{B}+3)\boldsymbol{K_{sat_j}} \left[\frac{0.5 \cdot (\theta_j + \theta_{j+1})}{0.5 \cdot (\boldsymbol{\theta_{sat_j}} + \boldsymbol{\theta_{sat_{j+1}}})}\right]^{2\boldsymbol{B}+2} \left(\frac{0.5}{\boldsymbol{\theta_{sat_j}}}\right) \quad \text{if } 1 \leq j \leq N-1 \tag{4.39}$$

$$= (2\boldsymbol{B}+3)\boldsymbol{K_{sat_j}} \left[\frac{\theta_j}{\boldsymbol{\theta_{sat_j}}}\right]^{2\boldsymbol{B}+2} \left(\frac{1}{\boldsymbol{\theta_{sat_j}}}\right) \quad \text{if } j = N \tag{4.40}$$

as the difference between the interface depth and the upper and lower node is per definition equal. If the effective porosity of either layer j or $j+1$ is less than $\boldsymbol{wimp}$, then $K_j = 0$ and $\frac{\partial K_j}{\partial \theta_j} = 0$.

Note that the interface depth is by definition exactly halfway in between two subsequent nodes and therefore soil moisture at both nodes contribute equally to the hydraulic conductivity and the flow through one interface. Two subsequent

interfaces are not situated equally far from a node point, but the flows through 2 subsequent interfaces are assumed to contribute equally to the calculation of soil moisture at a node. In figure 4.4 the calculation of the water flow is schematically presented.

After the determination of the surface runoff (SurfaceRunoff.F90) and the infiltration (Infiltration.F90), the water content in the soil layers is calculated using Darcy's equation (SoilWater.F90). Next, the soil moisture profile (w_{liq}, [kg .m^{-2}]) is updated for baseflow runoff and water in excess in the soil column (Drainage.F90). The baseflow drainage, R_b (qflx_drain, Drainage.F90), is obtained as the sum of the baseflows for unsaturated (qflx_drain_dry) and saturated (qflx_drain_wet) regions:

$$R_b = (1 - f_{sat}) \cdot \boldsymbol{K_d} \cdot \bar{w}_b^{(2 \cdot \boldsymbol{B}+3)} + f_{sat} \cdot \boldsymbol{l_d} \cdot \exp(-z_w) \quad [\text{kg.m}^{-2}.\text{s}^{-1}] \tag{4.41}$$

with

R_b	=	baseflow drainage [kg .m^{-2}.s^{-1}] (qflx_drain=qflx_drain_dry+qflx_drain_wet,Drainage.F90)
$\boldsymbol{K_d}$	=	saturated soil hydraulic conductivity at the bottom, parameter $\boldsymbol{kd}$ [mm.s^{-1}]
$\boldsymbol{l_d}$	=	base flow parameter for saturated fraction, parameter $\boldsymbol{ld}$ [mm.s^{-1}]
$\bar{w}_b$	=	weighted soil wetness for bottom $\boldsymbol{NwRb}$ layers [m.mm.s^{-1}]
	=	$\frac{\sum_{j=\boldsymbol{NwRb},N} s_j \cdot K_j \Delta z_j}{\sum_{j=\boldsymbol{NwRb},N} K_j \Delta z_j}$ (wsat,Drainage.F90)
$\boldsymbol{NwRb}$	=	first bottom layer contributing to the calculation of surface runoff [-], parameter $\boldsymbol{NwRb}$

The bottom soil layers (from $\boldsymbol{NwRb}$ through N) are updated by:

$$w_{liq,j} = w_{liq,j} - R_b \frac{\Delta z_j K_j}{\sum_{j=\boldsymbol{NwRb},N} K_j \Delta z_j} \cdot \Delta t \quad [\text{kg.m}^{-2}] \quad \text{for} \quad \boldsymbol{NwRb} < j < N-1 \tag{4.42}$$

All soil layers are finally updated by compensating shortages of water in one layer by moving water from the soil layer below. Shortage in the bottom layer is filled by water from R_b, resulting in R_b'. Water in excess of saturation, R_{excess} (xs/dtime, Drainage.F90) for each depth is added to the remaining R_b' and removed from the soil column. From the bottom layer, an outflux of $q_N = K_N$ is contributing to the total baseflow drainage, so that the total runoff, R_{tot} (qflx_drain, Drainage.F90), is given by:

$$R_{tot} = R_b + R_{excess} + K_N + \frac{\partial K_N}{\partial \theta_N} \Delta \theta_N \quad [\text{mm.s}^{-1}] \tag{4.43}$$

Water is allowed to pond on the surface soil layer so that the maximum amount of water for this layer is defined as

$$w_{liq,1} = \boldsymbol{w_{pond}} + \boldsymbol{\theta_{sat_1}} \Delta z_1 \tag{4.44}$$

with $\boldsymbol{w_{pond}}$ the ponding depth [mm] (pondmx, Drainage.F90), parameter $\boldsymbol{pondmx}$.

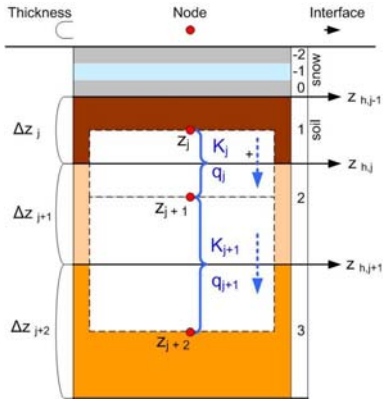


Figure 4.4: Soil water flow calculation in CLM.

Water balance equations (BalanceCheck.F90)

At time step n the water balance is checked through:

$$\begin{aligned} err_{H_2O} &= \{[w_{can} + w_{snow} + \sum_{j=1,N} (w_{liq,j} + w_{ice,j})]^n \\ &\quad -[w_{can} + w_{snow} + \sum_{j=1,N} (w_{liq,j} + w_{ice,j})]^{n-1} \\ &\quad -(P - (E_c + E_g) - R_s - R_b)\}\Delta t \end{aligned} \tag{4.45}$$

which should approach zero, with

w_{snow}	=	the mass of snow water [mm]
E_c	=	water flux from wet foliage [kg.m^{-2}.s^{-1}]
E_g	=	evaporation from ground [kg.m^{-2}.s^{-1}]

4.8.3 Heat fluxes and states

It is shown in the above section that evapotranspiration is an important factor in the determination of the soil water content. It forms the coupling between the water and the energy balance (Biogeophysics1.F90,Biogeophysics2.F90), which is subject of discussion in the following sections. Some variables related to the fluxes are schematically illustrated in figure 4.5.

Foliage fluxes (CanopyFluxes.F90)

Foliage energy conservation is given by the foliage energy budget equation (CanopyFluxes.F90):

$$\boxed{f(T_c) = R_{n,c} - H_c - \boldsymbol{L_v} E_c = 0 \ \ [\text{W.m}^{-2}]} \tag{4.46}$$

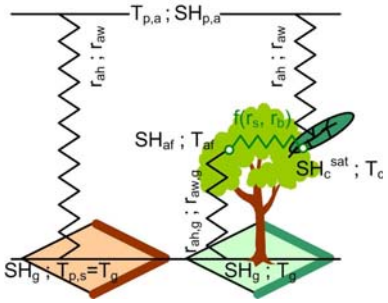


Figure 4.5: Resistances for heat fluxes over bare soil and vegetated surfaces. Differences in temperature cause sensible heat fluxes H, which are determined by r_{ah} and $r_{ah,g}$, while differences in specific humidity cause latent heat fluxes E, limited by r_{aw} and $r_{aw,g}$.

with

$R_{n,c}$	=	net radiation absorbed by canopy [W.m^{-2}]
	=	$S_{n,c} + L_{n,c}$ (net short- and longwave radiation, see further)
H_c	=	sensible heat flux from foliage to canopy air [W.m^{-2}] (eflx_sh_veg, efsh)
E_c	=	water flux from wet foliage [kg.m^{-2}.s^{-1}] (qflx_evap_veg), transpiration and evaporation
$\boldsymbol{L_v}$	=	coefficient of latent heat of evaporation [J.kg^{-1}]
T_c	=	leaf skin temperature [K] (t_veg)

A Newton-Raphson iteration is used to solve for the leaf skin temperature, T_c that balances the surface energy budget (CanopyFluxes.F90), because the vegetation temperature is coupled to the fluxes. This method typically searches for the roots of Eq. (4.46) based on the first terms of the Taylor expansion:

$$f(T_c^n) + \frac{df(T_c^n)}{dT_c} \cdot (T_c^{n+1} - T_c^n) = 0 \tag{4.47}$$

or after insertion of Eq. (4.46):

$$T_c^{n+1} - T_c^n = \frac{R_{n,c} - H_c - \boldsymbol{L_v} E_c}{-\frac{\partial R_{n,c}}{\partial T_c} + \frac{\partial H_c}{\partial T_c} + \boldsymbol{L_v}\frac{\partial E_c}{\partial T_c}} \tag{4.48}$$

Successive approximations of T_c^{n+1} are calculated until convergence, with n the iteration number. Initial guesses for $R_{n,c}$, H_c, $\boldsymbol{L_v} E_c$ and the partial derivatives are obtained by calculating the equations below with T_c^n equal to the value of previous time step or iteration step (tlbef). $R_{n,c}$, H_c and $\boldsymbol{L_v} E_c$ are updated for each T_c during the iteration.

Sensible heat fluxes

The air within the canopy has negligible heat capacity so that sensible heat fluxes are related by:

$$\boxed{H_a = H_c + H_g \ \ [\text{W.m}^{-2}]} \tag{4.49}$$

with

H_a = heat flux to the atmosphere [W.m^{-2}] (eflx_sh_tot, Biogeophysics2.F90, BareGroundFluxes.F90)

H_c = heat flux from the foliage [W.m^{-2}] (eflx_sh_veg, efsh, CanopyFluxes.F90)

H_g = heat flux from the ground to canopy space [W.m^{-2}] (eflx_sh_grnd, CanopyFluxes.F90, Biogeophysics2.F90, BareGroundFluxes.F90)

The equations for the heat fluxes are solved for the unknown canopy air temperature, T_{af} (CanopyFluxes.F90), using the ground temperature T_g of the previous time step and foliage temperature T_c of the previous time step or iteration step:

$$H_a = \rho_a \boldsymbol{c_p} \frac{(T_{af} - T_{p,a})}{r_{ah}} \quad [\text{W.m}^{-2}] \tag{4.50}$$

$$H_g = \rho_a \boldsymbol{c_p} \frac{(T_g - T_{af})}{r_{ah_g}} \quad [\text{W.m}^{-2}] \tag{4.51}$$

$$H_c = \rho_a \boldsymbol{c_p} (LAI + SAI) \frac{(T_c - T_{af})}{r_b} \quad [\text{W.m}^{-2}] \tag{4.52}$$

with

T_{af} = temperature of air within canopy [K], surface temperature at height $z_{0h} + d$

$T_{p,a}$ = air potential temperature at reference height z_r [K]

T_g = ground temperature, temperature of upper soil layer [K] (tg)

T_c = foliage temperature [K] (t_veg)

ρ_a = intrinsic density of air [kg.m^{-3}] (forc_rho), based on atmospheric forcings

$\boldsymbol{c_p}$ = specific heat capacity for dry air [J.kg^{-1}.K^{-1}], constant

r_{ah} = aerodynamic resistance for sensible heat [s.m^{-1}]

r_{ah_g} = thermal resistance to heat transfer [s.m^{-1}] between the ground and the canopy

r_b = leaf laminar boundary-layer resistance for vapor flux [s.m^{-1}]

- r_{ah}: the aerodynamic resistance for sensible heat [s.m^{-1}] between $z_{0h} + d$ (T_{af}) and z_r ($T_{p,a}$), in the presence of vegetation, is determined by Eq. (4.76), based on the Monin-Obukhov similarity theory (FrictionVelocity.F90, MoninObukIni.F90). See section on bare soil fluxes below, where the latent heat flux is exactly H_a defined above, except that T_{af} is replaced by T_g, the ground temperature. (rah(1), r_{ah}^{-1}, wta,SensibleHCond.F90)

- r_{ah_g}: the thermal resistance [s.m^{-1}] (rah(2), $r_{ah_g}^{-1}$, wtg, SensibleHCond.F90) to heat transfer between the ground (T_g) and the canopy height at $z_{0h} + d$ (T_{af}) is given by:

$$r_{ah_g} = (\boldsymbol{C_{soilc}} u_{af})^{-1} \tag{4.53}$$

with

$\boldsymbol{C_{soilc}}$ = transfer coefficient between canopy air and underlying ground, parameter $\boldsymbol{Csoilc}$ [-]
u_{af} = wind velocity within foliage layer [m.s^{-1}] (uaf,CanopyFluxes.F90)
= $V_a\sqrt{(\frac{1}{r_{am}V_a})} = u_*$ = friction velocity [m.s^{-1}]
V_a = wind speed including stability effect [m.s^{-1}]
r_{am} = aerodynamic resistance [s.m^{-1}]
= V_a/u_*^2

- r_b: the leaf laminar boundary-layer resistance for vapor flux [s.m^2.μ mol^{-1}], [s.m^{-1}] (rb, CanopyFluxes.F90, Stomata.F90, SensibleHCond.F90, $r_b^{-1}(LAI+SAI)$, wtl, SensibleHCond.F90) is given by:

$$r_b = \frac{1}{\boldsymbol{C_v}}\sqrt{\frac{\boldsymbol{D_f}}{u_{af}}} \tag{4.54}$$

with
$\boldsymbol{D_f}$ = characteristic leaf dimension, parameter $\boldsymbol{dleaf}$ [m]
$\boldsymbol{C_v}$ = turbulent transfer coefficient between canopy surface and air, constant = 0.01 [m.s$^{-1/2}$]

- $z_{0m}, z_{0h} = z_{0m,v}$, $z_{0h,v}$: the roughness lengths [m] for momentum and sensible heat are needed in the calculations of the resistances r_{am} and r_{ah} by the Monin-Obukhov similarity theory. They both equal $z_{0m,v}$ (z0m), which is in the presence of vegetation given by:

$$z_{0m,v} = \boldsymbol{R_{z0m}}.\boldsymbol{z_{top}} \tag{4.55}$$

with
$\boldsymbol{R_{z0m}}$ = ratio of momentum roughness to canopy top height pft-dependent parameter $\boldsymbol{z0mr}$ (z0mr) [-]
$\boldsymbol{z_{top}}$ = canopy height, surface parameter $\boldsymbol{MONTHLY\_HEIGHT\_TOP}$ [m] (htop)

- d: in the presence of vegetation, the displacement height is also needed in the calculations of the resistances r_{am} and r_{ah} and given by:

$$d = \boldsymbol{R_d}.\boldsymbol{z_{top}} \tag{4.56}$$

with
$\boldsymbol{R_d}$ = ratio of displacement height to canopy top height, parameter $\boldsymbol{displar}$ [-]

- z_r: the reference height [-] is given as input in the off-line version of CLM and corresponds to the height of the observations for the atmospheric forcings (forc_hgt).

Consequently, T_{af} (taf, CanopyFluxes.F90) is found by:

$$T_{af} = \frac{r_{ah}^{-1}T_{p,a} + (LAI+SAI)r_b^{-1}T_c + r_{ah_g}^{-1}T_g}{r_{ah}^{-1} + (LAI+SAI)r_b^{-1} + r_{ah_g}^{-1}} \tag{4.57}$$

Through substitution of Eq. (4.57) into Eq. (4.52), H_c (efsh) is calculated with T_c of previous time step.

Latent heat fluxes of water
Similar to the assumption above, the canopy air is assumed to have a negligible water storage, so that the (latent heat) fluxes of water are given by:

$$\boxed{E_a = E_c + E_g \ \ [\mathrm{kg.m^{-2}.s^{-1}}]} \tag{4.58}$$

with
E_a = evapotranpiration from canopy air to the atmosphere [$\mathrm{kg.m^{-2}.s^{-1}}$] (qflx_evap_tot, Biogeophysics2.F90)
E_c = evaporation and transpiration flux from dry and wet foliage
= $E_w + E_{tr}$ (qflx_evap_veg, CanopyFluxes.F90) [$\mathrm{kg.m^{-2}.s^{-1}}$]
E_g = evaporation from ground [$\mathrm{kg.m^{-2}.s^{-1}}$] (qflx_evap_soi, CanopyFluxes.F90, Biogeophysics2.F90)
All latent heat fluxes to the atmosphere are assumed to be positive.

The equations for the fluxes of water vapor are solved for the air specific humidity within the canopy, SH_{af}, with the value of the air specific humidity of ground surface air SH_g and the saturated specific humidity at the wet foliage SH_c^{sat} of the previous time step or iteration step:

$$E_a = \rho_a \frac{(SH_{af} - SH_{p,a})}{r_{aw}} \tag{4.59}$$

$$E_g = \rho_a \frac{(SH_g - SH_{af})}{r_{aw_g}} \tag{4.60}$$

$$E_c = r'' E_f^{pot} \tag{4.61}$$

$$E_f^{pot} = \rho_a (LAI + SAI) \frac{(SH_c^{sat} - SH_{af})}{r_b} \tag{4.62}$$

and

$$r'' = f_{wet} + f_{dry} \frac{r_b}{r_b + r_s} \quad \text{if } E_f^{pot} > 0 \text{ and } btran > 1 \cdot 10^{-10} \tag{4.63}$$

$$= f_{wet} \quad \text{if } E_f^{pot} > 0 \text{ and } btran \leq 1 \cdot 10^{-10} \tag{4.64}$$

$$= 1 \quad \text{if } E_f^{pot} \leq 0 \tag{4.65}$$

$$E_{tr} = f_{dry} \frac{r_b}{r_b + r_s} E_f^{pot} \quad \text{if } E_f^{pot} > 0 \text{ and } btran > 1 \cdot 10^{-10} \tag{4.66}$$

$$= 0 \quad \text{if } E_f^{pot} \leq 0 \text{ or } btran < 1 \cdot 10^{-10} \tag{4.67}$$

$$E_w = E_c - E_{tr} \tag{4.68}$$

with

SH_{af}	=	air specific humidity within canopy [kg.kg^{-1}] (qaf)
$SH_{p,a}$	=	lowest model level water vapor specific humidity [kg.kg^{-1}]
SH_c^{sat}	=	saturated specific humidity at wet foliage [kg.kg^{-1}](qsatl, Qsat.F90) at the foliage temperature T_c
SH_g	=	air specific humidity of ground surface air [kg.kg^{-1}](qg)
	=	$RH \cdot SH_g^{sat}$
RH	=	relative humidity [-] (hr, Biogeophysics1.F90)
	=	$\exp\left(\frac{\psi_1}{\boldsymbol{R_w} T_g}\right)$, with $\boldsymbol{R_w}$ the water vapor gas constant, ψ_1 negative potential of the upper soil layer
ρ_a	=	intrinsic density of air [kg.m^{-3}] (forc_rho), based on atmospheric forcings
E_f^{pot}	=	potential evaporation from wet foliage (leaves and stem) per unit wetted area [kg.m^{-2}.s^{-1}], potential latent heat flux (efpot, CanopyFluxes.F90)
E_c	=	vegetation evaporation (qflx_evap_veg,CanopyFluxes.F90) $E_w + E_{tr}$ [kg.m^{-2}.s^{-1}]
E_w	=	flux of water from wetted foliage [kg.m^{-2}.s^{-1}] (qflx_evap_veg-qflx_tran_veg, CanopyFluxes.F90)
E_{tr}	=	transpiration over dry leaf surface and only outward [kg.m^{-2}.s^{-1}] (qflx_tran_veg, CanopyFluxes.F90)
f_{wet}	=	wetted fraction of the canopy [−] (fwet, Hydrology1.F90)
f_{dry}	=	dry fraction of the canopy [−] (fdry, Hydrology1.F90)
r''	=	fraction of potential evaporation from leaf [-] (rpp, CanopyFluxes.F90)
r_{aw}	=	aerodynamic resistance for latent heat [s.m^{-1}]
r_{aw_g}	=	latent heat resistance between ground and canopy air [s.m^{-1}]
r_b	=	leaf laminar boundary-layer resistance for vapor flux [s.m^{-1}]
r_s	=	weighted stomatal resistance of sunlit and shaded foliage [s.m^{-1}]

- r_{aw}: the aerodynamic resistance for latent heat [s.m^{-1}] (raw(1), r_{aw}^{-1}, wtaq, LatentHCond.F90) between $z_{0w} + d$ (SH_{af}) and z_r ($SH_{p,a}$) is calculated similarly as r_{ah} (see next section), based on the Monin-Obukhov similarity theory (FrictionVelocity.F90, MoninObukIni.F90).

- r_{aw_g}: the latent heat resistance [s.m^{-1}] between ground (SH_g) and canopy air at height $z_{0w} + d$ (SH_{af}) (raw(2), $r_{aw_g}^{-1}$, wtgq, LatentHCond.F90) is given by:

$$r_{aw_g} = (\boldsymbol{C_{soilc}} u_{af})^{-1} \tag{4.69}$$

  with the conductance $r_{aw_g}^{-1}$ (wtgq, LatentHCond.F90) used in the calculations.

- r_b: the leaf laminar boundary-layer resistance for vapor flux [s.m^2.μ mol^{-1}], [s.m^{-1}] is calculated by Eq. (4.54). The expression $r_b^{-1}(LAI + SAI)r''$ (wtlq, LatentHCond.F90) is used in the calculations.

- r_s: the weighted stomatal resistance of sunlit and shaded foliage [s.m^{-1}] is determined as will be discussed in the section on the photosynthetic activity.

- $z_{0m,v}$, $z_{0q,v}$: the roughness lengths for momentum and latent heat are needed to calculate r_{am} and r_{aw}. They equal $z_{0m,v}$, given by Eq. (4.55) in the presence of vegetation.

Consequently, SH_{af} (qaf, CanopyFluxes.F90) is found by:

$$SH_{af} = \frac{r_{aw}^{-1} SH_{p,a} + r''(LAI + SAI) r_b^{-1} SH_c^{sat} + r_{aw_g}^{-1} SH_g}{(r_{aw}^{-1} + r''(LAI + SAI) r_b^{-1} + r_{aw_g}^{-1})} \tag{4.70}$$

Through substitution of Eq. (4.70) into Eq. (4.61), the latent heat flux of energy, $L_v H_c$ (efe) is calculated with SH_c^{sat} calculated based on T_c of previous time step (Qsat.F90).

Latent heat fluxes

The water fluxes [kg.m^{-2}.s^{-1}] (qflx-notation) are converted to latent heat fluxes (eflx-notation) by multiplication with a constant coefficient for latent heat of evaporation for water, $\boldsymbol{L_v}$ [J.kg^{-1}] (hvap), resulting in heat fluxes of evaporation $\boldsymbol{L_v} E_c$ from vegetation and $\boldsymbol{L_v} E_g$ from the ground surface [W.m^{-2}]. The total latent heat $\boldsymbol{L_v} E_{tot}$ is given by:

$$\boldsymbol{L_v} E_{tot} = \boldsymbol{L_v} E_c + \boldsymbol{L_v} E_g \quad [\text{W.m}^{-2}] \tag{4.71}$$

with

$\boldsymbol{L_v} E_c$	=	latent heat flux from leaves [W.m^{-2}] (hvap*qflx_evap_veg, Biogeophysics2.F90, efe, efsh, eflx_sh_veg, CanopyFluxes.F90)
$\boldsymbol{L_v} E_g$	=	latent heat flux from the ground (evaporation) [W.m^{-2}] (eflx_lh_grnd=htvp*qflx_evap_soi, Biogeophysics2.F90)
$\boldsymbol{L_v} E_{tot}$	=	total latent heat flux [W.m^{-2}] (eflx_lh_tot, Biogeophysics2.F90)

H_c and $\boldsymbol{L_v} E_c$ are used in Eq. (4.48), which yields an update of T_c. Next, E_f^{pot}, E_c, E_{tr}, and $\boldsymbol{L_v} E_c$ are updated and after the update of SH_c^{sat}, new values of T_{af} and SH_{af} are calculated and the variables used for the determination of the resistances are recalculated.

Summarized, it is through iteration (successive approximations) that leaf temperature T_c and consequently canopy air temperature, air specific humidity and resistance values within the canopy are found and that the fluxes can be calculated. During this iteration, the surface temperature T_g from a previous time step (CanopyFluxes.F90) is used. Eq. (4.46) is used to check the energy balance in the canopy. The surface temperature will be updated (SoilTemperature.F90) using the initially calculated heat fluxes resulting from the procedure described above. Finally (Biogeophysics2.F90), the heat fluxes will be recalculated for the updated surface temperature.

For the calculation of the soil temperature profile, the derivative of the soil heat fluxes (cgrnd) with respect to soil temperature [W.m^{-2}.K^{-1}] are needed. This derivative (CanopyFluxes.F90) is given by a addition of the derivatives of soil sensible

heat flux (cgrndl) and soil latent heat flux (cgrnds) with respect to soil temperature:

$$\frac{\partial H_g}{\partial T_g} + \boldsymbol{L_v}\frac{\partial E_g}{\partial T_g} = \rho_a \boldsymbol{c_p} {r_{ah_g}}^{-1} \frac{{r_{ah}}^{-1} + (LAI+SAI){r_b}^{-1}}{{r_{ah}}^{-1} + (LAI+SAI){r_b}^{-1} + r_{ah_g}^{-1}} + \boldsymbol{L_v}\rho_a {r_{aw_g}}^{-1} \frac{(r_{aw}^{-1} + r''(LAI+SAI)r_b^{-1})}{(r_{aw}^{-1} + r''(LAI+SAI)r_b^{-1} + {r_{aw_g}}^{-1})} \frac{\partial SH_g}{\partial T_g} \quad (4.72)$$

with $\frac{\partial SH_g}{\partial T_g}$ the temperature derivative of SH_g (dqgdT, qg, Biogeophysics1.F90).

If the updated evaporation (Biogeophysics2.F90) from the ground has negative values, the absolute value is taken to be dew, a flux towards the ground or snow (qflx_dew_ground, qflx_dew_snow). The ground surface evaporation rate $E_{g,Surf}$ (qflx_evap_grnd, Biogeophysics2.F90), used in the calculation of the infiltration, equals E_g (qflx_evap_soi) in case the evaporation is less than the amount of water in the upper soil/snow layer per time unit. If less water is stored in the upper layer than the amount to be evaporated, then $E_{g,Surf}$ is limited to that amount.

Bare soil fluxes (BareGroundFluxes.F90)

For the calculation of the heat fluxes over vegetation, the aerodynamic resistance for sensible and latent heat are calculated from integration of turbulent fluxes as for bare soil. When no vegetation is present, only these resistances are of importance. For bare soil or snow covered vegetation, the sensible and latent heat fluxes (H eflx_sh_tot=eflx_sh_grnd and E qflx_evap_tot=qflx_evap_soi) and the derivative of soil heat flux (cgrnd=cgrnds+htvp*cgrndl) with respect to the ground temperature T_g are respectively given by:

$$H = \rho_a \boldsymbol{c_p} \frac{T_{p,s} - T_{p,a}}{r_{ah}} \quad [\mathrm{W.m^{-2}}] \quad (4.73)$$

$$E = \rho_a \frac{SH_{p,s} - SH_{p,a}}{r_{aw}} \quad [\mathrm{kg.m^{-2}.s^{-1}}] \quad (4.74)$$

$$\frac{\partial H}{\partial T_g} + \boldsymbol{L_v}\frac{\partial E}{\partial T_g} = \frac{\rho_a \boldsymbol{c_p}}{r_{ah}} + \boldsymbol{L_v}\frac{\rho_a}{r_{aw}}\frac{\partial SH_{p,s}}{\partial T_g} \quad [\mathrm{W.m^{-2}.K^{-1}}] \quad (4.75)$$

with

$T_{p,s}$	=	ground surface temperature, defined at $z_{0h}+d$ [K] (tg)
	=	T_1, by definition = T_g for bare soil only
$T_{p,a}$	=	atmospheric potential temperature [K]
	=	$T_a + 0.0098 z_r$ (forc_th,thm=forc_t+0.0098[K.m^{-1}]forc_hgt_t)
$SH_{p,s}$	=	specific humidity at the ground surface, defined at $z_{0w}+d$
	=	SH_g for bare soil only (qg) [kg.kg^{-1}]
$SH_{p,a}$	=	atmospheric specific humidity (forc_q) [kg.kg^{-1}]
$\frac{\partial SH_g}{\partial T_g}$	=	the temperature derivative of SH_g (dqgdT, Biogeophysics1.F90)
r_{ah}	=	aerodynamic resistances for sensible heat transfer [s.m^{-1}]
r_{aw}	=	aerodynamic resistances for latent heat transfer [s.m^{-1}]

- r_{ah}, r_{aw}: the aerodynamic resistances r_{ah} and r_{aw} for sensible and latent heat transfer (raih, raiw, BareGroundFluxes.F90, rah(1), CanopyFluxes.F90) between the atmosphere at height z_r and the surface at height $z_{0h}+d$ are respectively given by:

$$r_{ah} = \frac{T_{p,a} - T_{p,s}}{u_* T_*} \tag{4.76}$$

$$r_{aw} = \frac{SH_{p,a} - SH_{p,s}}{u_* SH_*} \tag{4.77}$$

with
u_* = velocity scale, friction velocity [m.s^{-1}] (ustar,FrictionVelocity.F90)
T_* = temperature scale [K] (tstar,BareGroundFluxes.F90,CanopyFluxes.F90)
SH_* = moisture scale [kg.kg^{-1}] (qstar,BareGroundFluxes.F90,CanopyFluxes.F90)

- u_*, T_*, SH_*: the scales based on the the Monin-Obukhov similarity theory, derived from the wind profile u_*/V_a (V_a [m.s^{-1}] is the wind speed including the stability effect), potential temperature profile $\frac{T_{v*}}{T_{p,a}-T_{p,s}}$ (temp1, FrictionVelocity.F90), and potential humidity profile $\frac{SH_{v*}}{SH_{p,a}-SH_{p,s}}$ (temp2, FrictionVelocity.F90), respectively. These profiles all depend on the Monin-Obukhov length $\mathscr{L}$ [Garratt, 1992; Pahlow *et al.*, 2001], several constants and roughness lengths. Calculation of u_* includes the calculation of momentum fluxes and the aerodynamic resistance for momentum [s.m^{-1}] between $z_{0m}+d$ and z_r.

- $z_{0m,g}, z_{0h,g}, z_{0w,g}$: the roughness lengths over ground [m] for momentum, sensible heat and water vapor (z0mg, z0hg, z0qg,Biogeophysics1.F90, BareGroundFluxes.F90) are needed in the calculation of the wind, temperature and humidity profile and assumed to be all equal to the roughness length for soil ***zlnd*** (zlnd) [m], in case no snow is covering the soil, or equal to the roughness length for snow ***zsno*** (zsno) [m], in case of snow. z_{0h} and z_{0w} are identical and a function of u_* and $z_{0m} =$ ***zlnd*** in case of bare soil.

$$z_{0h,g} = z_{0w,g} = z_{0m,g} \exp\left[-0.13\left(\frac{u_* z_{0m,g}}{15 \cdot 10^{-5}}\right)^{0.45}\right] \tag{4.78}$$

- d : zero-plane displacement height is 0 m for bare soil [m] (displa)

For the calculation of the heat fluxes from bare soil as well as from vegetated patches, integration of the turbulent flux profiles is necessary to determine aerodynamic resistances (r_{am}, r_{ah}, r_{aw}). This is achieved through an initialization of the Monin-Obukhov length ($\mathscr{L}$), followed by the effective calculation of the friction velocity (u_*) and the relation for potential temperature and humidity profiles ($\frac{T_{v*}}{T_{p,a}-T_{p,s}}$ and $\frac{SH_{v*}}{SH_{p,a}-SH_{p,s}}$) of the surface boundary layer. The first guesses are adapted for by 2 more iterations, until the aerodynamic resistances are correctly approached. Notice that in case of bare soil, photosynthesis should not be considered.

Soil temperature (SoilTemperature.F90)

To perform the calculations described in the section above, the ground surface temperature is needed to derive the net absorbed radiation, the evaporation from the ground and the latent heat flux from the ground.

Using a Crank-Nicholson scheme, the integrated energy balance equation over soil layer j with thickness Δz_j, can be numerically expanded as:

$$\boxed{c_j \frac{(T_j^{n+1} - T_j^n)\Delta z_j}{\Delta t} = \boldsymbol{\omega}[F_j^n - F_{j-1}^n] + (1-\boldsymbol{\omega})[F_j^{n+1} - F_{j-1}^{n+1}] \ \ [\mathrm{W.m^{-2}}]} \tag{4.79}$$

with

T_j	=	temperature in layer j [K] (t_soisno, SoilTemperature.F90)
F_j	=	heat flux across the interface between layer j and $j+1$ [W.m^{-2}], positive upward (fn, SoilTemperature.F90)
c_j	=	volumetric heat capacity [J.m^{-3}.K^{-1}]
$\boldsymbol{\omega}$	=	weighting coefficient or relaxation factor in time domain, parameter ***cnfac*** [-]
Δz_j	=	$z_{h,j} - z_{h,j-1}$
$(.)^n$	=	at the beginning of the time step
$(.)^{n+1}$	=	at the end of the time step
Δt	=	1 time step [s]

For the surface, interior and bottom layer, the heat fluxes (schematically shown in figure 4.6) are respectively defined by:

$$F_{j-1} = \boldsymbol{\omega} F_{j-1}^n + (1-\boldsymbol{\omega}) F_{j-1}^{n+1} \tag{4.80}$$

$$= -h^{n+1} = -\left[h^n + \frac{\partial h}{\partial T_j} \cdot (T_j^{n+1} - T_j^n)\right] \tag{4.81}$$

$$F_j = \lambda(z_{h,j}) \frac{T_{j+1} - T_j}{z_{j+1} - z_j} \tag{4.82}$$

$$F_N = 0 \tag{4.83}$$

with

h	=	net incoming heat flux, positive downward [W.m^{-2}]
$\lambda(z_{h,j})$	=	thermal conductivity at interface depth $z_{h,j}$ [W.m^{-1}.K^{-1}]

- h : the incoming energy in the upper surface layer (hs, SoilTemperature.F90) is positive downward and given by

$$h = R_{n,g} - H_g - \boldsymbol{L_v} E_g \ \ [\mathrm{W.m^{-2}}] \tag{4.84}$$

with

$R_{n,g}$	=	net radiation reaching the ground [W.m^{-2}]
	=	$S_{n,g} + L_{d,g} + (1 - f_{vns})\boldsymbol{\epsilon}_g L_{d,a} - L_{u,g}$ (see later)

The derivative $\frac{\partial R_{n,g}}{\partial T}$ (dhsdT=-cgrnd-4.*emg * sb * tg**3) is calculated by:

$$\frac{\partial h}{\partial T_j} = -\frac{\partial L_{u,g}}{\partial T_j} - \frac{\partial H_g}{\partial T_j} - \boldsymbol{L_v}\frac{\partial E_g}{\partial T_j} \quad [\mathrm{W.m^{-2}.K^{-1}}] \tag{4.85}$$

using Eq. (4.72) and (4.75) for $\frac{\partial H_g}{\partial T_j}$ and $\frac{\partial E_g}{\partial T_j}$ (CanopyFluxes.F90, BareGroundFluxes.F90). The radiation terms $R_{n,g}$, $L_{u,g}$ and the derivative will be further explained in next section.

- c_j : the volumetric heat capacity ($c_j \Delta z$=cv(i), SoilTermProp.F90) is given by:

$$c_j = \sum_{k=d,liq,ice} \boldsymbol{\rho_k c_k} \theta_k \quad [\mathrm{J.m^{-3}.K^{-1}}] \tag{4.86}$$

with
$(\boldsymbol{\rho_d c_d})_j$ = specific heat capacity of dry soil in layer j [J.m^{-3}.K^{-1}], parameter $\boldsymbol{csol_j}$ (csol(j))
$(\boldsymbol{\rho_k c_k})_j$ = specific heat capacity of water and ice [J.m^{-3}.K^{-1}], constants, for $k = liq$ and $k = ice$ respectively (cpice,cpliq)

- $\lambda(z_{h,j})$: the thermal conductivity at the interface $z_{h,j}$ (tk(i), SoilTermProp.F90) is function of the thermal conductivities of the surrounding layers j and $j+1$ and given by:

$$\lambda(z_{h,j}) = \frac{\lambda_j \lambda_{j+1}(z_{j+1} - z_j)}{\lambda_j(z_{j+1} - z_{h,j}) + \lambda_{j+1}(z_{h,j} - z_j)} \quad [\mathrm{W.m^{-1}.K^{-1}}] \quad \text{for } j = 1, \ldots, N-1 \tag{4.87}$$

$$= 0 \quad [\mathrm{W.m^{-1}.K^{-1}}] \quad \text{for } j = N \tag{4.88}$$

with
λ_j = thermal conductivity for layer j (at node depth z_j)
= $K_e(\boldsymbol{\lambda_{sat}} - \boldsymbol{\lambda_{dry}}) + \boldsymbol{\lambda_{dry}}$ [W.m^{-1}.K^{-1}] (tk(i), thk(i))
$\boldsymbol{\lambda_{sat}}$ = thermal conductivity for saturated soil [W.m^{-1}.K^{-1}], in case of unfrozen soil, parameter $\boldsymbol{tksatu_j}$ (dksat,tksatu(i)), or in case of frozen soil, dependent on thermal conductivity for soil minerals, parameter $\boldsymbol{tkmg}$ [W.m^{-1}.K^{-1}]
$\boldsymbol{\lambda_{dry}}$ = thermal conductivity for dry natural soil, parameter $\boldsymbol{tkdry_j}$ [W.m^{-1}.K^{-1}] (tkdry(i)), which depends on the bulk density parameter $\boldsymbol{bd}$ [kg.m^{-3}]
K_e = Kersten number [-]
= $\log_{10} \frac{(\theta_{ice}+\theta_{liq})}{(1-\theta_d)} + 1.$ (dke, SoilTermProp.F90)
$(1-\theta_d)$ = porosity, with θ_d the fraction dry soil [-]

Eq. (4.87) is based on the assumption that the flux across the interface is equal to that from the node j to the interface and the flux from the interface to the node $j+1$.

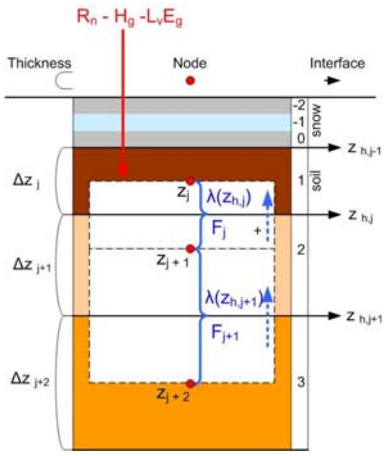


Figure 4.6: Heat fluxes in a soil/snow column in CLM.

This yields for Eq. (4.79):

$$\begin{aligned}
T_j^{n+1} &= T_j^n + \left\{\boldsymbol{\omega} F_j^n + (1-\boldsymbol{\omega}) F_j^{n+1}\right\} \frac{\Delta t}{c_j \Delta z_j} \\
&\quad - - \left\{h + \frac{\partial h}{\partial T_j} \cdot (T_j^{n+1} - T_j^n)\right\} \frac{\Delta t}{c_j \Delta z_j} \qquad (4.89) \\
&= T_j^n + \left\{\boldsymbol{\omega}\left[\lambda_{j*}(T_{j+1}^n - T_j^n)\right] + (1-\boldsymbol{\omega})\left[\lambda_{j*}(T_{j+1}^{n+1} - T_j^{n+1})\right]\right\} \frac{\Delta t}{c_j \Delta z_j} \\
&\quad - - \left\{h + \frac{\partial h}{\partial T_j} \cdot (T_j^{n+1} - T_j^n)\right\} \frac{\Delta t}{c_j \Delta z_j} \qquad (4.90) \\
T_j^{n+1} &= T_j^n + \left\{\boldsymbol{\omega}\left[\lambda_{j*}(T_{j+1}^n - T_j^n) - \lambda_{j-1*}(T_j^n - T_{j-1}^n)\right]\right\} \frac{\Delta t}{c_j \Delta z_j} \\
&\quad + \left\{(1-\boldsymbol{\omega})\left[\lambda_{j*}(T_{j+1}^{n+1} - T_j^{n+1}) - \lambda_{j-1*}(T_j^{n+1} - T_{j-1}^{n+1})\right]\right\} \frac{\Delta t}{c_j \Delta z_j} \qquad (4.91) \\
T_j^{n+1} &= T_j^n + \left\{\boldsymbol{\omega}\left[0 - \lambda_{j-1*}(T_j^n - T_{j-1}^n)\right]\right\} \frac{\Delta t}{c_j \Delta z_j} \\
&\quad + \left\{(1-\boldsymbol{\omega})\left[0 - \lambda_{j-1*}(T_j^{n+1} - T_{j-1}^{n+1})\right]\right\} \frac{\Delta t}{c_j \Delta z_j} \qquad (4.92)
\end{aligned}$$

for a surface, interior and bottom layer, respectively, with

$\lambda_{j*} = \frac{\lambda(z_{h,j})}{z_{j+1} - z_j}$ (tk(j)/dzp, tk(j)/dzm)

Eq. (4.79) can then be written as a tridiagonal system (Tridiagonal.F90) for soil layer temperature as:

$$a_j T_{j-1}^{n+1} + b_j T_j^{n+1} + c_j T_{j+1}^{n+1} = r_j \qquad (4.93)$$

with for the surface layer ($j = 1$):

$$\begin{aligned} a_j &= 0 \\ b_j &= 1 + \left[(1-\boldsymbol{\omega})\lambda_{j*} - \frac{\partial h}{\partial T_j}\right]\frac{\Delta t}{c_j \Delta z_j} \\ c_j &= -(1-\boldsymbol{\omega})\lambda_{j*}\frac{\Delta t}{c_j \Delta z_j} \\ r_j &= T_j^n + \left[h - \frac{\partial h}{\partial T_j}\cdot T_j^n + \boldsymbol{\omega} F_j^n\right]\frac{\Delta t}{c_j \Delta z_j} \end{aligned}$$

for the interior layers ($1 < j < N$):

$$\begin{aligned} a_j &= -(1-\boldsymbol{\omega})\lambda_{j-1*}\frac{\Delta t}{c_j \Delta z_j} \\ b_j &= 1 + (1-\boldsymbol{\omega})(\lambda_{j*} + \lambda_{j-1*})\frac{\Delta t}{c_j \Delta z_j} \\ c_j &= -(1-\boldsymbol{\omega})\lambda_{j*}\frac{\Delta t}{c_j \Delta z_j} \\ r_j &= T_j^n + \boldsymbol{\omega}[F_j^n - F_{j-1}^n]\frac{\Delta t}{c_j \Delta z_j} \end{aligned}$$

and for the bottom layer ($j = N$):

$$\begin{aligned} a_j &= -(1-\boldsymbol{\omega})\lambda_{j-1*}\frac{\Delta t}{c_j \Delta z_j} \\ b_j &= 1 + (1-\boldsymbol{\omega})\lambda_{j-1*}\frac{\Delta t}{c_j \Delta z_j} \\ c_j &= 0 \\ r_j &= T_j^n - \boldsymbol{\omega} F_{j-1}^n \frac{\Delta t}{c_j \Delta z_j} \end{aligned}$$

with a_j, b_j, c_j and r_j given by at(j), bt(j), ct(j), rt(j).

The surface layer temperature calculated this way is averaged over the surface layer and hence differs from the surface skin temperature. Therefore, to provide an accurate surface temperature T_g, the heat capacity is tuned through adjustment of the layer thickness by:

$$\Delta z_{j*} = 0.5(z_j - z_{h,j-1} + \boldsymbol{c_a}(z_{j+1} - z_{h,j-1})) \quad \text{for} \quad j = 1 \tag{4.94}$$

with

$\boldsymbol{c_a}$ = tuning factor to turn first layer temperature T_1 into surface temperature T_g [-], parameter ***capr***

The upper soil layer temperature is assigned to the ground temperature (tg = t_soisno(snl+1)).

Depending on the resulting temperature, a phase change may occur and adaptations in sensible heat and temperatures can be calculated accordingly (PhaseChange.F90).

Radiation absorbed by canopy and ground

The radiation is an important input for the calculations of heat fluxes (see above). The solar radiation absorbed by vegetation $S_{n,c}$ is an important factor in the process of photosynthesis and the solar radiation absorbed by ground $S_{n,g}$ together with the downward longwave radiation below the canopy $L_{d,g}$ form a boundary condition for the calculation of the temperature profile in the soil.

Solar fluxes
The solar shortwave radiation (diffuse + direct (d/i), visible + near infrared (λ_v/λ_n, vis/nir)) is partially reflected and partially absorbed by the surface (vegetation and ground), by the vegetation and by the ground. The net absorbed radiations [W.m^{-2}] are given by (SurfaceRadiation.F90):

$$S_n = S_{n,c} + S_{n,g} \tag{4.95}$$

$$S_{n,c} = \sum_{\lambda_v/\lambda_n} \sum_{d/i} \phi ab_{\lambda_v/\lambda_n,d/i} S_{\lambda_v/\lambda_n,d/i} \quad \text{if vegetated} \tag{4.96}$$

$$= 0 \ \text{W.m}^{-2} \quad \text{if not vegetated} \tag{4.97}$$

$$S_{n,g} = \sum_{\lambda_v/\lambda_n} \{(1-\bar{\alpha}_{\lambda_v/\lambda_n,d(g)})(\phi tr_{dd} S_{\lambda_v/\lambda_n,d}) + (1-\bar{\alpha}_{\lambda_v/\lambda_n,i(g)}) [(\phi tr_{ii} S_{\lambda_v/\lambda_n,i}) + (\phi tr_{id} S_{\lambda_v/\lambda_n,d})]\} \quad \text{if vegetated} \tag{4.98}$$

$$= \sum_{\lambda_v/\lambda_n} \{(1-\bar{\alpha}_{\lambda_v/\lambda_n,d(g)})(\phi tr_{dd} S_{\lambda_v/\lambda_n,d}) + (1-\bar{\alpha}_{\lambda_v/\lambda_n,i(g)})[(\phi tr_{ii} S_{\lambda_v/\lambda_n,i})]\} \quad \text{if not vegetated} \tag{4.99}$$

with

$S_{n,c}$	=	solar radiation absorbed by vegetation [W.m^{-2}] (sabv)
$S_{n,g}$	=	solar radiation absorbed by ground [W.m^{-2}] (sabg)
S_n	=	total solar radiation absorbed [W.m^{-2}] (fsa)
$S_{\lambda_v/\lambda_n,d/i}$	=	incident solar radiation [W.m^{-2}] (forc_solad(ib),forc_solai(ib),atmdrvMod.F90)
$\bar{\alpha}_{\lambda_v/\lambda_n,d/i(g)}$	=	effective ground (bare soil) albedo [-] (albgri,albgrd)
$\phi ab_{\lambda_v/\lambda_n,d/i}$	=	flux absorbed by vegetation [-] per unit direct/diffuse flux (fabd, fabi, SurfaceAlbedo.F90)
$\phi tr_{d/i\|d/i}$	=	transmitted down direct/diffuse flux below vegetation [-] per unit direct/diffuse flux (ftdd, ftii,ftii(ib),SurfaceAlbedo.F90), e.g., ϕtr_{id} is the direct transmitted flux per unit diffuse flux. Note that ϕtr_{di} is always 0. $\phi tr_{dd} = 1$ without vegetation, and dependent on optical pft-parameters otherwise.

The reflected amount of solar radiation (fsr, SurfaceRadiation.F90) is given by:

$$S_r = \sum_{d/i} \overline{\alpha_{\lambda_v,d/i}} S_{\lambda_v,d/i} + \sum_{d/i} \overline{\alpha_{\lambda_n,d/i}} S_{\lambda_n,d/i} \quad [\mathrm{W.m^{-2}}] \tag{4.100}$$

with
$\overline{\alpha_{\lambda_v/\lambda_n,d/i}}$ = weighted (over bare soil and vegetation) surface albedo [-],
= flux reflected above the vegetation layer per unit incoming flux,
with $\overline{\alpha_{\lambda_v,d/i}}$ (albd(1),albi(1),SurfaceAlbedo.F90)
and $\overline{\alpha_{\lambda_n,d/i}}$ (albd(2),albi(2),SurfaceAlbedo.F90,fre,TwoStream.F90)
for vis and nir radiation, respectively

Note that the 2 terms in the above addition can be used to compute the normalized difference vegetation index (NDVI).

- $\alpha_{\lambda_v/\lambda_n,d/i(g)}$: the albedo for bare soil (albgrd, albgri, SoilAlbedo.F90, SurfaceAlbedo.F90) is calculated by:

$$\alpha_{\lambda_v/\lambda_n,d/i(g)} = \min[\boldsymbol{\alpha_{sat}(\lambda_v/\lambda_n)} + \Delta\alpha(\theta_{liq,1}), \boldsymbol{\alpha_{dry}(\lambda_v/\lambda_n)}] \tag{4.101}$$

with
$\Delta\alpha_g(\theta_{liq,1})$ = increase in albedo due to dryness of the surface soil [-]
= $\max[0.01 \cdot (11 - 40\theta_{liq,1}), 0]$ (inc, SoilAlbedo.F90)
$\boldsymbol{\alpha_{sat/dry}}$ = albedo for saturated/dry soil [-], dependent on solor class through parameter ***SOIL_COLOR***, with $\boldsymbol{\alpha_{sat/dry}(\lambda_n)} = 2 \cdot \boldsymbol{\alpha_{sat/dry}(\lambda_v)}$

The effective albedo over bare soil $\bar{\alpha}_{\lambda_v/\lambda_n,d/i(g)}$ (albgrd(ib), albgri(ib), SoilAlbedo.F90) takes into account the possibility of snow cover, but is equal to $\alpha_{\lambda_v/\lambda_n,d/i(g)}$ otherwise.

- $\alpha_{\lambda_v/\lambda_n,d/i(c)}$: the albedo over vegetation (TwoStream.F90, SurfaceAlbedo.F90) is calculated as function (f) of features of a two-stream approximation model [Bonan, 1996]:

$$\alpha_{\lambda_v/\lambda_n,d/i(c)} = f(\mu, LAI, SAI, \boldsymbol{taul/s}_{\lambda_v/\lambda_n}, \boldsymbol{rhol/s}_{\lambda_v/\lambda_n}, \boldsymbol{xl}) \tag{4.102}$$

with
μ = cosine of solar zenith angle [-],
depends on the latitude and sun position
$\boldsymbol{taul/s}_{\lambda_v/\lambda_n}$ = pft-parameters, leaf/stem transmittance [-]
$\boldsymbol{rhol/s}_{\lambda_v/\lambda_n}$ = pft-parameters, leaf/stem reflectance [-]
for vis or nir radiation, respectively
$\boldsymbol{xl}$ = leaf/stem orientation index [-] (xl,chil,TwoStream.F90)
pft-parameter $\boldsymbol{xl}$

The effective albedo over vegetation $\bar{\alpha}_{\lambda_v/\lambda_n,d/i(c)}$ takes into account the possibility of snow cover, but is equal to $\alpha_{\lambda_v/\lambda_n,d/i(c)}$ otherwise.

The albedo over vegetation is implicitly used to determine the downward direct ($\phi_{tr_{dd}}$, ftd, Twostream.F90, ftdd) and diffuse ($\phi_{tr_{id}}$, $\phi_{tr_{ii}}$, fti, Twostream.F90, ftid, ftii) fluxes below vegetation and the reflected ($\overline{\alpha_{\lambda_v,d/i}}$, $\overline{\alpha_{\lambda_n,d/i}}$, fre, Twostream.F90, albd, albi) and absorbed (ϕ_{ab}, fab, Twostream.F90, fabd, fabi) flux per unit incoming radiation.

- $\overline{\alpha_{\lambda_v/\lambda_n,d/i}}$: the weighted albedo over a grid cell is given by (albd, albi, SurfaceAlbedo.F90):

$$\overline{\alpha_{\lambda_v/\lambda_n,d/i}} = (1 - F_{veg})\bar{\alpha}_{\lambda_v/\lambda_n,d/i(g)} + F_{veg}\bar{\alpha}_{\lambda_v/\lambda_n,d/i(c)} \tag{4.103}$$

with
F_{veg} = fraction of vegetation [-]

Longwave fluxes
Downward atmospheric longwave radiation $L_{d,a}$ is absorbed by the canopy (CanopyFluxes.F90, Biogeophysics2.F90) and the ground through:

$$\begin{aligned}
L_{n,c} &= f_{vns}(\epsilon_v[1 + (1 - \epsilon_v)(1 - \boldsymbol{\epsilon}_g)]L_{d,a} \\
&\quad -\epsilon_v[2 - \epsilon_v(1 - \boldsymbol{\epsilon}_g)]\boldsymbol{\sigma}T_c^4 + \epsilon_v L_{u,g}) && (4.104)\\
L_{n,g} &= L_{d,g} - L_{u,g} && (4.105)\\
L_{d,g} &= \boldsymbol{\epsilon}_g[(1 - f_{vns}) + f_{vns}(1 - \epsilon_v)]L_{d,a} + f_{vns}\boldsymbol{\sigma}\epsilon_v\boldsymbol{\epsilon}_g T_c^4 && (4.106)\\
L_{u,g} &= \boldsymbol{\sigma}\boldsymbol{\epsilon}_g T_g^4 && (4.107)\\
L_{u,c} &= f_{vns}[(1 - \boldsymbol{\epsilon}_g)(1 - \epsilon_v)(1 - \epsilon_v)L_{d,a} \\
&\quad +\epsilon_v[1 + (1 - \boldsymbol{\epsilon}_g)(1 - \epsilon_v)]\boldsymbol{\sigma}T_c^4 \\
&\quad +(1 - \epsilon_v)L_{u,g}] && (4.108)\\
L_u &= L_{u,c} + (1 - f_{vns})[(1 - \boldsymbol{\epsilon}_g)L_{d,a} + L_{u,g}] && (4.109)\\
L_{n,u} &= L_u - L_{d,a} && (4.110)
\end{aligned}$$

with

ϵ_v	=	emissivity of the vegetation, $= 1 - \exp[-(LAI + SAI)]$ [-] (emv, Biogeophysics1.F90)
$\boldsymbol{\epsilon}_g$	=	emissivity of the ground, $= 0.96$ [-] (emg, Biogeophysics1.F90)
$\boldsymbol{\sigma}$	=	Stefan-Boltzmann constant (sb), $= 5.67 \cdot 10^{-8}$ [W.m^{-2}.K^{-4}]
f_{vns}	=	fraction of vegetation not buried with snow [-] (frac_veg_nosno)
	=	$(1 - F_{snow})F_{veg}$, with F_{snow} and F_{veg} the fraction of snow and vegetation
$L_{d,a}$	=	downward atmospheric longwave radiation [W.m^{-2}] (forc_lwrad)
$L_{n,c}$	=	net longwave radiation absorbed by vegetation [W.m^{-2}]
$L_{n,g}$	=	net longwave radiation absorbed by ground [W.m^{-2}]
$L_{d,g}$	=	downward longwave radiation below the canopy [W.m^{-2}]
$L_{u,g}$	=	upward longwave radiation from ground [W.m^{-2}]
$L_{u,c}$	=	upward longwave radiation above the canopy [W.m^{-2}]
L_u	=	outgoing long-wave radiation from vegetation and ground [W.m^{-2}]
$L_{n,u}$	=	net upward longwave radiation [W.m^{-2}]

The different fluxes are indicated in figure 4.2. The calculations for longwave radiation is spread over the modules CanopyFluxes.F90 ($L_{n,c}$ used in calculation of T_c, $L_{d,g}$ = dlrad, $L_{u,c}$ = ulrad) and Biogeophysics2.F90 (L_u = eflx_lwrad_out, $L_{n,u}$ = eflx_lwrad_net).

The radiative skin temperature of the surface is an important diagnostic variable, e.g. in remote sensing analysis. It is defined from the outgoing longwave flux from the surface, L_u through (t_rad, Biogeophysics2.F90):

$$T_{rad} = \left(\frac{L_u}{\boldsymbol{\sigma}}\right)^{1/4} \quad [\mathrm{K}] \tag{4.111}$$

Recall that the boundary condition for soil heat fluxes at the top of a soil column is given by:

$$h = R_{n,g} - H_g - L_v E_g \tag{4.112}$$

with
$R_{n,g}$ = $S_{n,g} + L_{d,g} + (1 - f_{vns})\boldsymbol{\epsilon}_g L_{d,a} - L_{u,g}$ [W.m^{-2}]
h = ground heat flux [W.m^{-2}], positive directed towards the soil (eflx_soil_grnd, Biogeophysics2.F90, hs, SoilTemperature.F90)

The derivative of $R_{n,g}$ is given by:

$$\frac{\partial L_{u,g}}{\partial T_j} = 4\boldsymbol{\epsilon}_g \sigma T_g^3 \tag{4.113}$$

Photosynthesis, transpiration and stomatal resistance (Stomata.F90, CanopyFluxes.F90)

In case vegetation is present, the stomatal resistance, r_s (rs, Stomata.F90, rssun, rssha, CanopyFluxes.F90) is a factor needed in the calculation of the transpiration (flux of latent heat). In Eq. (4.63) and (4.66) the term $\frac{r_b}{r_b+r_s}$ appears, which is in the code calculated as:

$$r_b \frac{\left(\frac{LAI_{sun}}{r_b+r_{s,sun}} + \frac{LAI_{sha}}{r_b+r_{s,sha}}\right)}{LAI + SAI} \tag{4.114}$$

with
r_b = boundary layer resistance [s.m^{-1}]
r_s = leaf stomatal resistance [s.m^{-1}]

For use in calculations on photosynthetic processes, the units [s.m^{-1}] are converted to [s.m^2 μ mol^{-1}] by division through $cf = 10^6 p_s \cdot [0.001\boldsymbol{R}(T_a + 0.0098 z_r)]^{-1}$, with p_s the atmospheric pressure [Pa] and $\boldsymbol{R}$ the universal gas constant.

The stomatal resistance for sunlit and shaded leaves are calculated by solving the equations for the photosynthesis in this section, using average absorbed photosynthetically active radiation (PAR) for sunlit and shaded leaves, yielding photosynthesis A_{sun} and A_{sha} [μ mol CO$_2$.m^{-2}.s^{-1}] (psnsun, psnsha).

The photosynthetically active radiation (PAR) for sunlit (PAR_{sun}) and shaded leaves (PAR_{sha}) [W.m^{-2}] is calculated as (parsun, parsha, SurfaceRadiation.F90, compare to Eq. (4.96)):

$$PAR_{sun} = \sum_{d/i} \phi_{ab_{v,d/i}} S_{\lambda_v,d/i} \frac{LAI}{LAI + SAI} \quad (4.115)$$

$$PAR_{sha} = 0 \text{ W.m}^{-2} \quad (4.116)$$

and the total photosynthesis A [μ mol CO$_2$.m^{-2}.s^{-1}] (fpsn, Biogeophysics1.F90) is given by the combination of photosynthesis for sunlit leaves and shaded leaves:

$$A = LAI_{sun}\, A_{sun} + LAI_{sha}\, A_{sha} \quad (4.117)$$

with

LAI_{sun}	=	sunlit leaf area [-], $LAI \cdot f_{sun}$ (laisun, SurfaceRadiation.F90)
LAI_{sha}	=	shaded leaf area [-], $LAI \cdot f_{sha}$ (laisha, SurfaceRadiation.F90)
f_{sun}	=	sunlit fraction of canopy [-] (fsun, SurfaceAlbedo.F90) $\frac{1}{k_b(LAI+SAI)}(1 - \exp[-k_b(LAI + SAI)])$, if $k_b > 1.10^{-6}$
f_{sha}	=	$1 - f_{sun}$ [-] (fsha)
k_b	=	PAR extinction coefficient for direct beam solar radiation [-], dependent on parameters describing the radiative transfer [Bonan, 1996] within vegetative canopies based on the two-stream approximation, such as the pft-parameters $\boldsymbol{rhol}_{\lambda_v}$ and $\boldsymbol{rhos}_{\lambda_v}$ [-] for leaf/stem reflectance, $\boldsymbol{taul}_{\lambda_v}$ and $\boldsymbol{taus}_{\lambda_v}$ [-] for leaf/stem transmittance, and $\boldsymbol{xl}$ [-] a leaf/stem orientation index

The stomatal resistance, r_s [s.m^2.μ mol^{-1}], is coupled to the photosynthesis A [μ mol CO$_2$.m^{-2}.s^{-1}] (psn) through equation:

$$\frac{1}{r_s} = \boldsymbol{m_p} \frac{A\, e_s}{CO_{2,s}\, e_i} p_s + \boldsymbol{b} \quad (4.118)$$

with

p_s	=	atmospheric pressure at the surface (input) [Pa] (forc_pbot)
e_s	=	vapor pressure at the leaf surface [Pa]
e_i	=	vapor pressure inside leaf, saturated vapor pressure at T_c [Pa] (ei)
$CO_{2,s}$	=	CO$_2$ concentration at the leaf surface [Pa] (cs)
$\boldsymbol{m_p}$	=	slope of the conductance-to-photosynthesis relationship, pft-dependent parameter $\boldsymbol{mp}$ [-]
$\boldsymbol{b}$	=	minimal stomatal conductance when A=0 μ mol.m^{-2}.s^{-1} (bp=2000,Stomata.F90)

The leaf photosynthesis A [μ mol CO$_2$.m^{-2}.s^{-1}] (psn, Stomata.F90) is determined by the most limiting factor for carboxylation. $A = \min(w_c, w_j, w_e)$ (wc, wj, we,Stomata.F90), with w_c (Eq. (4.119)) the RuBP carboxylase (Rubisco) limited carboxylation rate, w_j (Eq. (4.120)) the light limited rate (or maximum capacity to generate RuBP),

and w_e (Eq. (4.121)) the export limited rate for C_3 plants or the PEP carboxylase limited rate for C_4 plants:

$$w_c = \frac{(CO_{2,i} - \Gamma_*)V_{max}}{CO_{2,i} + K_c(1 + \frac{O_{2,i}}{K_o})}\boldsymbol{C_{3/4}} + V_{max}(1 - \boldsymbol{C_{3/4}}) \quad (4.119)$$

$$w_j = \frac{(CO_{2,i} - \Gamma_*)4.6\, PAR\, \boldsymbol{q_{25}}}{CO_{2,i} + 2\Gamma_*}\boldsymbol{C_{3/4}} + 4.6\, PAR\, \boldsymbol{q_{25}}(1 - \boldsymbol{C_{3/4}}) \quad (4.120)$$

$$\text{for } \; CO_{2,i} - \Gamma_* \geq 0$$

$$w_e = \frac{V_{max}}{2}\boldsymbol{C_{3/4}} + \frac{4000\, V_{max}\, CO_{2,i}}{p_s}(1 - \boldsymbol{C_{3/4}}) \quad (4.121)$$

with

$\boldsymbol{C_{3/4}}$	=	photosynthetic pathway, for $C_3 = 1$, for $C_4 = 0$, pft-dependent parameter ***c3psn*** [-]
$CO_{2,i}$	=	internal CO_2 concentration in leaf [Pa] (ci), updated by $\max(CO_{2,s} - 1.65Ap_sr_s, 0)$ during iterative calculation of A
$O_{2,i}$	=	internal O_2 concentration [Pa] (o2,also in CanopyFluxes.F90)
	=	$0.209p_s$
PAR	=	absorbed photosynthetically active radiation [W.m^{-2}] (apar)
	=	converted to photon flux by 4.6μ mol photon.J^{-1} (ppf)
Γ_*	=	CO_2 compensation point [Pa] (cp),
	=	$\frac{1}{2}\frac{K_c}{K_o}0.21O_{2,i}$
K_c	=	Michaelis-Menten constant for CO_2 [Pa] (kc),
	=	$\boldsymbol{K_{c25}Q_{10,c}^{(T_c-273.16-25)/10}}$
K_o	=	Michaelis-Menten constant for O_2 [Pa] (ko),
	=	$\boldsymbol{K_{o25}Q_{10,o}^{(T_c-273.16-25)/10}}$
$\boldsymbol{K_{c25}}$	=	CO_2 Michaelis-Menten constant at 25 ℃ [Pa] (kc25=30)
$\boldsymbol{K_{o25}}$	=	O_2 Michaelis-Menten constant at 25 ℃ [Pa] (ko25=30000)
$\boldsymbol{Q_{10,c}}$	=	Q_{10} value for $\boldsymbol{K_{c25}}$ (temperature sensitivity) [-] (akc=2.1)
$\boldsymbol{Q_{10,o}}$	=	Q_{10} value for $\boldsymbol{K_{o25}}$ (temperature sensitivity) [-] (ako=1.2)
$\boldsymbol{q_{25}}$	=	quantum efficiency at 25℃[μ mol CO_2. μ mol photon^{-1}], pft-dependent parameter ***qe25*** (qe25)
V_{max}	=	maximum carboxylation rate [μ mol CO_2.m^{-2}. s^{-1}]
	=	$\boldsymbol{V_{max,25}Q_{10,v}^{(T_c-273.16-25)/10}}f_T^{-1}(T_c - 273.16)f_w(w)$
	=	(vcmx,CanopyFluxes.F90)
$\boldsymbol{V_{max,25}}$	=	maximum carboxylation rate at 25℃[μ mol CO_2.m^{-2}. s^{-1}], pft-dependent parameter ***vcmx25***
$\boldsymbol{Q_{10,v}}$	=	Q_{10} value for $\boldsymbol{V_{max,25}}$ (temperature sensitivity constant) [-] (avcmx=2.4)

f_* = functions to represent variation of the maximum rate of carboxylation with temperature (*=T) and soil water (*=w) (dependency on nitrogen is incorporated in $\boldsymbol{V_{max,25}}$)
f_T = $\left\{1+\exp\left[\frac{-220000+710(T_c)}{\boldsymbol{R}T_c}\right]\right\}$ [-]
$\boldsymbol{R}$ = universal gas constant = 8.314 J.K^{-1}.mol^{-1}
f_w = soil water transpiration factor [-] (btran), decreases photosynthesis and increases stomatal resistance as soil dries,
= $btran$, see section on soil water

In order to solve Eq. (4.118) for r_s, it should be noted that e_s can be written as function of r_s. The transpiration fluxes through the leaf and the surrounding atmosphere are related by:

$$\frac{e'_a - e_i}{r_b + r_s} = \frac{e'_a - e_s}{r_b} = \frac{e_s - e_i}{r_s} \tag{4.122}$$

and consequently:

$$e_s = \frac{e'_a r_b + e_i r_b}{r_b + r_s} \tag{4.123}$$

with
e'_a = constrained vapor pressure of air to prevent numerical instability [Pa]
= $\max[0.25e_i\boldsymbol{C_{3/4}} + 0.4e_i(1-\boldsymbol{C_{3/4}}), \min(e_a, e_i)]$ (cea,Stomata.F90)
e_a = vapor pressure of canopy air [Pa] (ea)
= $\frac{p_s SH_{af}}{0.622}$
r_b = leaf boundary resistance [s.m^2.μ mol^{-1}] (converted from [s.m^{-1}])

This serial relationship can also be found for the photosynthesis, which helps to determine $CO_{2,s}$ and $CO_{2,i}$:

$$A = \frac{CO_{2,a} - CO_{2,i}}{(1.37\,r_b + 1.65\,r_s)p_s} = \frac{CO_{2,a} - CO_{2,s}}{1.37\,r_b p_s} = \frac{CO_{2,s} - CO_{2,i}}{1.65\,r_s p_s} \tag{4.124}$$

and consequently $CO_{2,s}$ (cs,Stomata.F90) and $CO_{2,i}$ are given by:

$$CO_{2,s} = \max(CO_{2,a} - 1.37Ap_s r_b, 10^{-6}) \tag{4.125}$$
$$CO_{2,i} = \max(CO_{2,s} - 1.65Ap_s r_s, 0) \tag{4.126}$$

with
$CO_{2,a}$ = CO_2 concentration in the air [Pa]
= $355\cdot10^{-6}p_s$ (co2,CanopyFluxes.F90)

- During night time, the photosynthesis A=0 μ mol CO_2.m^{-2}.s^{-1} and the stomatal resistance r_s is assigned a constant value of $1/\boldsymbol{b}$ [s.m^2.μ mol^{-1}] (see Eq. (4.118)) and is used in the calculation of the latent heat fluxes after conversion to [s.m^{-1}].
- During day time, the stomatal resistance is found as the larger value of the 2 roots of the equation found by combining Eq. (4.118) and Eq. (4.123):

$$\boxed{\left(\frac{\boldsymbol{m_p} A p_s e'_a}{CO_{2,s} e_i} + \boldsymbol{b}\right) r_s^2 + \left(\frac{\boldsymbol{m_p} A p_s r_b}{CO_{2,s}} + \boldsymbol{b} r_b - 1\right) r_s - r_b = 0} \tag{4.127}$$

This equation is iterated three times with an initial value of

$$CO_{2,i} = 0.7\, CO_{2,a}\, \boldsymbol{C_{3/4}} + 0.4\, CO_{2,a}\, (1 - \boldsymbol{C_{3/4}}) \tag{4.128}$$

Once the initial $CO_{2,i}$ is known, A and $CO_{2,s}$ are calculated and r_s is determined. This value of r_s is used to compute an updated value for $CO_{2,i}$ by Eq. (4.126).

Checking of energy balance equations (BalanceCheck.F90)

The energy balance equations consist of several equations, of the error in the foliage (Eq. (4.129)) (Canopyfluxes.F90), solar radiation energy (Eq. (4.130)), longwave radiation energy (Eq. (4.131)) and surface energy (Eq. (4.132)), respectively :

$$\begin{aligned}
err_{Can} &= S_{n,c} + L_{n,c} - H_c - L_v E_c && (4.129)\\
err_S &= S_n + S_r - (S_{\lambda_v,d} + S_{\lambda_n,d} + S_{\lambda_v,i} + S_{\lambda_n,i}) && (4.130)\\
err_L &= L_u - L_{nu} - L_{d,a} && (4.131)\\
err_{Surf} &= S_{n,c} + S_{n,g} + L_{d,a} - L_u - (H_c + H_g) \\
&\quad -(L_v E_c + L_v E_g) - h && (4.132)
\end{aligned}$$

These errors should closely approach zero.

4.9 Summary

A wide variety of computer models can be used to represent natural systems. These models should be calibrated for each application through parameter optimization. The adapted structure of the physically based land surface model CLM2.0 was introduced, with special indication of all the parameters (bold fonts). Some default constant values have been turned into parameters, because the modeled soil moisture was found to be sensitive to them. The CLM2.0 will be used in next chapters to represent the OPE^3 field processes.

List of symbols for CLM2.0

$\alpha_{sat/dry}$	albedo for saturated/dry soil [-] dependent on color class
ϵ_g	emissivity of the ground, = 0.96 [-]
λ_{dry}	thermal conductivity for dry natural soil, parameter $tkdry_j$ [W.m^{-1}.K^{-1}]
λ_{sat}	thermal conductivity for saturated soil, parameter $tksatu_j$ [W.m^{-1}.K^{-1}]
θ_{sat_j}	volumetric soil water at saturation in layer j [m^3.m^{-3}]
ρ_k	density of constituent k [kg.m^{-3}]
$(\rho_{liq}c_{liq})_j$	specific heat capacity of water, constants [J.m^{-3}.K^{-1}]
$(\rho_{ice}c_{ice})_j$	specific heat capacity of ice, constants [J.m^{-3}.K^{-1}]
$(\rho_d c_d)_j$	specific heat capacity of dry soil in layer j [J.m^{-3}.K^{-1}], parameter $csol_j$
σ	Stefan-Boltzmann constant, = $5.67 \cdot 10^{-8}$ [W.m^{-2}.K^{-4}]
ψ_{max}	maximum value of negative of leaves potential before dessication or wilting point potential [mm]
ψ_{sat}	saturated soil water potential [mm]
ψ_{sat}	saturated soil water potential [mm]
ω	weighting coefficient in time domain, parameter $clm\%cnfac$ [-]
B	parameter defined by Clapp and Hornberger [1978] [-]
b	minimal stomatal conductance when A=0 [μmol.m^{-2}.s^{-1}]
c_a	tuning factor to turn first layer T into surface T [-], parameter $capr$
c_k	specific heat for constituent k [J.kg^{-1}.K^{-1}], constant
c_p	specific heat for dry air [J.kg^{-1}.K^{-1}], constant
$C_{3/4}$	photosynthetic pathway, for C_3 = 1, for C_4 = 0, parameter $c3psn$ [-]
C_{soilc}	transfer coefficient between canopy air and underlying ground, parameter $Csoilc$ [-]
D_f	characteristic leaf dimension, parameter $dleaf$ [m]
$dewmx$	maximum dew [mm]
f_z	water table depth scale parameter [m^{-1}]
K_d	saturated soil hydraulic conductivity at the bottom, parameter kd♣ [mm.s^{-1}]
K_{sat_0}	saturated hydraulic conductivity at the surface [mm.s^{-1}]
K_{sat_j}	saturated hydraulic conductivity at layer j [mm.s^{-1}]
K_{c25}	CO_2 Michaelis-Menten constant at 25 ℃ [Pa]
K_{o25}	O_2 Michaelis-Menten constant at 25 ℃ [Pa]
L_v	latent heat of evaporation [J.kg^{-1}], constant
l_d	base flow parameter for saturated fraction, parameter ld♣ [mm.s^{-1}]

To be continued ...

Continued ...	
$\boldsymbol{m_p}$	slope of the conductance-to-photosynthesis relationship, parameter $\boldsymbol{mp}$ [-]
$\boldsymbol{NwRb}$	first bottom layer contributing to the calculation of surface runoff [-], parameter $\boldsymbol{NwRb}$
$\boldsymbol{NwRs}$	last top layer contributing to the calculation of surface runoff [-]
$\boldsymbol{Q_{10,c}}$	Q_{10} value for $\boldsymbol{K_{c25}}$ (temperature sensitivity constant) [-]
$\boldsymbol{Q_{10,o}}$	Q_{10} value (temperature sensitivity constant) for $\boldsymbol{K_{o25}}$ [-]
$\boldsymbol{Q_{10,v}}$	Q_{10} value (temperature sensitivity constant) for $\boldsymbol{V_{max,25}}$ [-]
$\boldsymbol{q_{25}}$	quantum efficiency at 25℃, parameter $\boldsymbol{qe25}$ [μ mol CO_2. μ mol photon^{-1}]
$\boldsymbol{R}$	universal gas constant = 8.314 J.K^{-1}.mol^{-1}
$\boldsymbol{R_w}$	water vapor gas constant = 8.314/18.016 J.K^{-1}.kg^{-1}
$\boldsymbol{R_d}$	ratio of displacement height to canopy top height, $\sim 2/3$, parameter $\boldsymbol{displar}$
$\boldsymbol{R_{z0m}}$	ratio of momentum roughness to canopy top height, pft-dependent parameter $\boldsymbol{z0mr}$
$\boldsymbol{rhol}/\boldsymbol{s}_{\lambda_v/\lambda_n}$	pft-parameters, leaf/stem reflectance for vis or nir radiation [-]
$\boldsymbol{taul}/\boldsymbol{s}_{\lambda_v/\lambda_n}$	pft-parameters, leaf/stem transmittance for vis or nir radiation [-]
$\boldsymbol{V_{max,25}}$	maximum carboxylation rate at 25℃, parameter $\boldsymbol{vcmx25}$, [μ mol CO_2.m^{-2}. s^{-1}]
$\boldsymbol{w_{fact}}$	determined by the distribution of the topographic index [-]
$\boldsymbol{xl}$	leaf/stem orientation index [-], pft-parameter $\boldsymbol{xl}$
$\boldsymbol{z_*}$	length scale for K_{sat_j} decrease [−]
$\alpha_{\lambda_v/\lambda_n,d/i(c)}$	albedo of thick canopy [-], function of parameters $\boldsymbol{rhol}$, $\boldsymbol{tau}$, SAI, LAI, μ
$\alpha_{\lambda_v/\lambda_n,d/i(g)}$	albedo for bare soil [-], function of parameter $\boldsymbol{\alpha_{sat/dry}}$
$\bar{\alpha}_{\lambda_v/\lambda_n,d/i(c)}$	effective albedo over vegetation [-]
$\bar{\alpha}_{\lambda_v/\lambda_n,d/i(g)}$	effective ground (bare soil) albedo [-]
$\overline{\alpha_{\lambda_v/\lambda_n,d/i}}$	weighted (over bare soil and vegetation) surface albedo over a grid cell [-]
Γ_*	CO_2 compensation point [Pa]
$\Delta\alpha_g(\theta_{liq,1})$	increase in albedo due to dryness of the surface soil (layer $j = 1$)
$\Delta\theta_j^{n+1}$	$(\theta_j^{n+1} - \theta_j^n)$ [-]
Δz_j	thickness of soil layer j [m], $z_{h,j} - z_{h,j-1}$, with $z_{h,j}$ the interface depths
ϵ_v	emissivity of the vegetation, $= 1-\exp(-(LAI + SAI))$ [-]
$\theta_{ice,j}$	volumetric soil ice mass in layer j [m^3.m^{-3}]
$\theta_{liq,j}$	volumetric soil moisture in layer j [m^3.m^{-3}]
θ_j	$= \theta_{liq,j}$, if no confusion possible [m^3.m^{-3}]
θ_k	partial volume of constituent k [fraction]
$\lambda(z_{h,j})$	thermal conductivity at interface depth $z_{h,j}$ [W.m^{-1}.K^{-1}]
λ	thermal conductivity of medium [W.m^{-1}.K^{-1}]

To be continued ...

Continued ...	
λ_v/λ_n	vis/nir, radiation type
μ	cosine solar zenith angle [-]
ρ_a	intrinsic density of air [kg.m^{-3}]
$\phi ab_{\lambda_v/\lambda_n,d/i}$	flux absorbed by vegetation per unit direct/diffuse flux [-]
$\phi tr_{d/i\|d/i}$	transmitted down direct/diffuse flux below vegetation per unit direct/diffuse flux [-]
ψ_j	matrix potential of soil at layer j [mm]
A	total photosynthesis [μmol CO$_2$.m^{-2}.s^{-1}]
A_{sha}	photosynthesis for shaded leaves [μmol CO$_2$.m^{-2}.s^{-1}]
A_{sun}	photosynthesis for sunlit leaves [μmol CO$_2$.m^{-2}.s^{-1}]
c_j	volumetric heat capacity [J.m^{-3}.K^{-1}]
$CO_{2,a}$	CO$_2$ concentration in the air [Pa]
$CO_{2,i}$	internal CO$_2$ concentration in leaf [Pa]
$CO_{2,s}$	CO$_2$ concentration at the leaf surface [Pa]
d	zero-plane displacement height, proportional to canopy height
d/i	direct/diffuse
D_d	direct throughfall [mm.s^{-1}]
D_r	drainage of water from foliage and stem [mm.s^{-1}]
e'_a	constrained vapor pressure of air to prevent numerical instability [Pa]
e_a	vapor pressure of canopy air [Pa]
e_i	vapor pressure inside leaf, saturated vapor pressure at T_c [Pa]
e_s	vapor pressure at the leaf surface [Pa]
E	water flux [kg.m^{-2}.s^{-1}]
E_a	evapotranpiration from canopy air to the atmosphere [kg.m^{-2}.s^{-1}]
E_c	water flux (transpiration and evaporation) from wet foliage [kg.m^{-2}.s^{-1}]
E_f^{pot}	potential evaporation from wet foliage (leaves and stem) per unit wetted area, potential latent energy flux [kg.m^{-2}.s^{-1}]
$E_{g,Surf}$	soil surface evaporation [mm.s^{-1}]
E_g	evaporation from ground [kg.m^{-2}.s^{-1}]
E_{tot}	total latent heat flux of water [kg.m^{-2}.s^{-1}]
E_{tr}	transpiration from dry foliage [kg.m^{-2}.s^{-1}]
E_w	flux of water from wetted foliage [mm.s^{-1}]
f_*	functions to represent variation of the maximum rate of carboxylation [-]
$f_{r,j}$	root fraction within soil layer j [-]
$f_{root,j}$	effective root fraction in soil layer j [-]
$f_{root,j}E_{tr}$	water extracted by plant roots, transpiration [kg.m^{-2}.s^{-1}]
f_{sat}	partial contributing area, saturated fraction [-]
f_{sha}	shaded fraction of canopy [-]
f_{sun}	sunlit fraction of canopy [-]

To be continued ...

Continued ...	
f_T	variation of the maximum rate of carboxylation due to temperature [-]
f_w	soil water transpiration factor [-], decreases photosynthesis and increases stomatal resistance as soil dries
f_{vns}	fraction of vegetation excluding the part buried by snow [-]
F_{veg}	fraction of vegetation [-]
F_j	heat flux across the interface between layer j and $j+1$, positive upward [W.m^{-2}]
G_w	sum of effective precipitation and snowmelt [kg .m^{-2}.s^{-1}]
h	net incoming energy flux, positive downward [W.m^{-2}]
H	sensible heat flux [W.m^{-2}]
H_a	heat flux to the atmosphere [W.m^{-2}]
H_c	sensible heat flux from foliage to canopy air [W.m^{-2}]
H_g	heat flux from the ground to canopy space [W.m^{-2}]
k	*ice* (ice), *liq* (liquid water), v (vapor), d (dry soil) [J.kg^{-1}]
k_b	PAR extinction coefficient for direct beam solar radiation [-]
K_c	Michaelis-Menten constant for CO_2 [Pa]
K_e	Kersten number, factor depending on the degree of saturation [-]
K_j	unsaturated hydraulic conductivity at layer j [mm.s^{-1}]
K_o	Michaelis-Menten constant for O_2 [Pa]
$\mathscr{L}$	Monin-Obukhov length [m]
$L_{d,a}$	downward atmospheric longwave radiation [W.m^{-2}]
$L_{d,g}$	downward longwave radiation below the canopy [W.m^{-2}]
$L_{n,c}$	net longwave radiation absorbed by vegetation [W.m^{-2}]
$L_{n,g}$	net longwave radiation absorbed by ground [W.m^{-2}]
$L_{n,u}$	net upward longwave radiation [W.m^{-2}]
$L_{u,c}$	upward longwave radiation above the canopy [W.m^{-2}]
$L_{u,g}$	upward longwave radiation from ground [W.m^{-2}]
L_d	dry fraction of the canopy [−]
L_u	outgoing long-wave radiation from vegetation and ground [W.m^{-2}]
LAI	one-sided leaf area index with burying by snow [-]
LAI_{sha}	shaded leaf area [-]
LAI_{sun}	sunlit leaf area [-]
$M_{ice,liq}$	rate of phase change (ice to liquid) [kg.m^{-3}.s^{-1}]
$O_{2,i}$	$0.209p_s$, internal O_2 concentration [Pa]
p_s	atmospheric pressure at the surface (input) [Pa]
P	precipitation for 1 time step [mm.s^{-1}]
PAR	absorbed photosynthetically active radiation [W.m^{-2}]
PAR_{sha}	photosynthetically active radiation for shaded leaves [W.m^{-2}]
PAR_{sun}	photosynthetically active radiation for sunlit leaves [W.m^{-2}]
q_j	water flow between neighboring layers, positive downward [kg.m^{-2}.s^{-1}]

To be continued ...

Continued ...	
r''	fraction of potential evaporation from leaf [-]
r_{ah_g}	thermal resistance [s.m^{-1}]
r_{ah}	aerodynamic resistance for sensible heat [s.m^{-1}]
r_{aw_g}	latent heat resistance between ground and bottom of canopy [s.m^{-1}]
r_{aw}	aerodynamic resistance for latent heat [s.m^{-1}]
r_b	leaf laminar boundary-layer resistance for vapor flux [s.m^2.μ mol^{-1}],[s.m^{-1}]
r_s	weighted leaf stomatal resistance of sunlit and shaded foliage [s.m^{-1}]
R_b	baseflow drainage [kg .m^{-2}.s^{-1}]
R_{excess}	runoff from water in excess of saturation [mm.s^{-1}]
$R_{n,c}$	net radiation absorbed by canopy [W.m^{-2}]
$R_{n,g}$	net radiation reaching the ground [W.m^{-2}]
R_n	net radiation absorbed [W.m^{-2}]
R_s	surface runoff [kg .m^{-2}.s^{-1}]
R_{tot}	total runoff [mm.s^{-1}]
s_j	soil wetness (including ice) [-]
SAI	one-sided stem area index with burying by snow [-]
$S_{\lambda_v/\lambda_n,d/i}$	incident solar radiation [W.m^{-2}]
S_d	extrapolates soil water dependence of ground evaporation, currently not used in the code and set to 0
$S_{n,c}$	solar radiation absorbed by vegetation [W.m^{-2}]
$S_{n,g}$	solar radiation absorbed by ground [W.m^{-2}]
S_n	total solar radiation absorbed [W.m^{-2}]
S_k	source or sink term [kg.m^{-3}.s^{-1}]
S_m	snowmelt [kg .m^{-2}.s^{-1}]
S_r	reflected amount of solar radiation [W.m^{-2}]
SH_*	moisture scale [kg.kg^{-1}]
SH_{af}	air specific humidity within canopy [kg.kg^{-1}]
SH_c^{sat}	saturated specific humidity at the wet foliage [kg.kg^{-1}]
SH_g	air specific humidity of ground surface air [kg.kg^{-1}]
$SH_{p,a}$	lowest model level water vapor specific humidity [kg.kg^{-1}]
$SH_{p,s}$	specific humidity at the ground surface [kg.kg^{-1}]
T_*	temperature scale [K]
T_{af}	temperature of air within canopy [K]
$T_{p,a}$	atmospheric potential temperature [K]
$T_{p,s}$	ground surface temperature [K]
T_{rad}	radiative skin temperature of the surface [K]
T_c	leaf skin temperature [K]
T_g	ground temperature, temperature of upper soil layer [K]
T_j	temperature in layer j [K]
u_*	friction velocity [m.s^{-1}]

To be continued ...

Continued ...	
u_{af}	wind velocity within foliage layer [m.s^{-1}]
V_{max}	maximum carboxylation rate [μ mol CO_2.m^{-2}. s^{-1}]
w_{can}	canopy interception water store [mm]
w_{cmax}	maximum water capacity of the canopy [mm]
w_c	RuBP carboxylase (Rubisco) limited carboxylation rate of photosynthesis [μ mol CO_2.m^{-2}.s^{-1}]
w_e	export limited rate for C_3 plants or the PEP carboxylase limited rate for C_4 plants of photosynthesis [μ mol CO_2.m^{-2}.s^{-1}]
$w_{ice,j}$	ice mass in layer j [kg .m^{-2}=mm]
w_j	light limited rate (or maximum capacity to generate RuBP) of photosynthesis [μ mol CO_2.m^{-2}.s^{-1}]
$w_{liq,j}$	liquid water mass in layer j [kg .m^{-2}=mm]
$w_{LT}(j)$	transpiration restricting factor [-]
$\bar{w}_b$	weighted soil wetness for bottom $\boldsymbol{NwRb}$ layers [m.mm.s^{-1}]
$\bar{w}_s$	soil layer thickness weighted soil wetness in top $\boldsymbol{NwRs}$ layers [m]
$z_{0h,v/g}$	roughness lengths over vegetation/ ground for sensible heat [m]
$z_{0m,v/g}$	roughness lengths over vegetation/ ground for momentum [m]
$z_{0w,v/g}$	roughness lengths over vegetation/ ground for water vapor [m]
$z_{h,j}$	depth of interface of soil/snow layer j and $j+1$ [m]
z_j	node depth of layer j [m]
z_r	reference height, lowest model level, input [m]
z_w	mean water table depth [m]

Chapter 5

CLM2.0 calibration and initialization

5.1 Introduction

In chapter 4, the physical processes underlying the structure of CLM2.0 were described, with special indication of the parameters. Given the observed atmospheric forcings, deterministic runs with this model completely depend on the choice of the initial conditions and the parameters. In order to assure good simulations for the OPE^3 field by the CLM2.0, the parameters and the initial states were simultaneously estimated for all observed soil profiles. A correct system representation is needed to be able to rely on the dynamical constraints imposed by the model during the process of recursive state estimation in later chapters.

5.2 Linearized state transition matrix

To investigate the dependence between state variables (model structure), the system or transition matrix was determined for the linearized version of CLM2.0, using default tuning parameter values and empirical expressions for parameters in CLM2.0. Spatially averaged surface data (soil and vegetation) as well as meteorological forcings were taken from the OPE^3 field, which was assumed to cover 1 pixel in this test. Initial conditions were set to default values and assumed to have realistic values after 1 year of spin-up (2001). During 2002, for every first day of each month the system matrix was determined by changing each state variable separately and recording the impact of each change on all state variables at the next time step. A state transition matrix was calculated for the linearized system version as obtained through Taylor's expansion (see section 2.3.1), i.e. the matrix

is given by:

$$\mathbf{F}_{i+1,i} = \begin{bmatrix} \frac{\partial \mathrm{f}_1}{\partial x_1} & \cdots & \frac{\partial \mathrm{f}_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{f}_n}{\partial x_1} & \cdots & \frac{\partial \mathrm{f}_n}{x_n} \end{bmatrix} |(\mathbf{x}_i, \mathbf{u}_i) \tag{5.1}$$

with $\mathrm{f}_n = \mathrm{f}_{n,i+1,i}(x_{1,i}, \ldots, x_{n,i}, u_{1,i}, \ldots, u_{p,i}) = x_{n,i+1}$. The resulting state transition matrices for 12 months in 2002 are shown in figure 5.1. The state variables presented include vegetation temperature (x_1), vegetation water (x_2), soil temperature (x_3 through x_{12}) and soil moisture (x_{13} through x_{22}) at 10 depths. The 'initial' (time step i) temperatures were perturbed by 0.5 K and liquid water variables by 0.01 kg.m^{-2} to find the resulting change in all state variables $x_{a,i+1}$, with $a = 1, \ldots, n$. The first and 3th through 12th column give $\frac{\Delta x_{a,i+1}}{0.5 \text{ K}}$, while the second and 13th through 22th columns give $\frac{\Delta x_{a,i+1}}{0.01 \text{ kg.m}^{-2}}$.

Vegetation temperature (x_1) was affected by vegetation water and by the upper soil temperature and only in May and April some effect from the upper soil moisture was found. The surface or upper soil temperature (x_3) was influenced by the temperature in maximum 5 deeper layers and by the water content in the 5 upper soil layers, maximally. Soil temperature in all layers is influenced by soil temperature and water content in surrounding layers. In the growing season, soil temperature (x_1, x_3 through x_7) was also affected by the canopy water content. Soil moisture (x_{13} through x_{22}) was of course affected by the soil moisture of surrounding layers and during spring almost all layers experience some influence from the canopy water and the upper soil temperature, probably through transpiration. During spring and summer months, soil temperature in deeper layers had an influence on soil moisture. For colder the months, soil temperature from the upper layers only had an influence on soil moisture. The fact that soil temperature and moisture were related is normal, and was found in field experiments like the one by Lakshmi *et al.* [2003]. The figure also illustrates that the influence of soil moisture in one layer on other layers is variable in time and that changes in the deeper layers have almost no effect on other soil layers.

Knowledge of the linearized system matrix is of direct importance in data assimilation techniques like KF and EKF. The system matrix also gives an indication of how corrections by assimilations will be propagated. Further, this matrix is crucial to analyze the system, its sensitivity to variations in different variables, its stability, and features like observability and controllability.

5.3 Practical input-output structure

The off-line version of CLM2.0 requires meteorological forcings as input, and land surface data and initial state estimates for parameter and state initialization. By default, output is written in NetCDF (binary) files. In this research only hourly soil moisture output values were investigated.

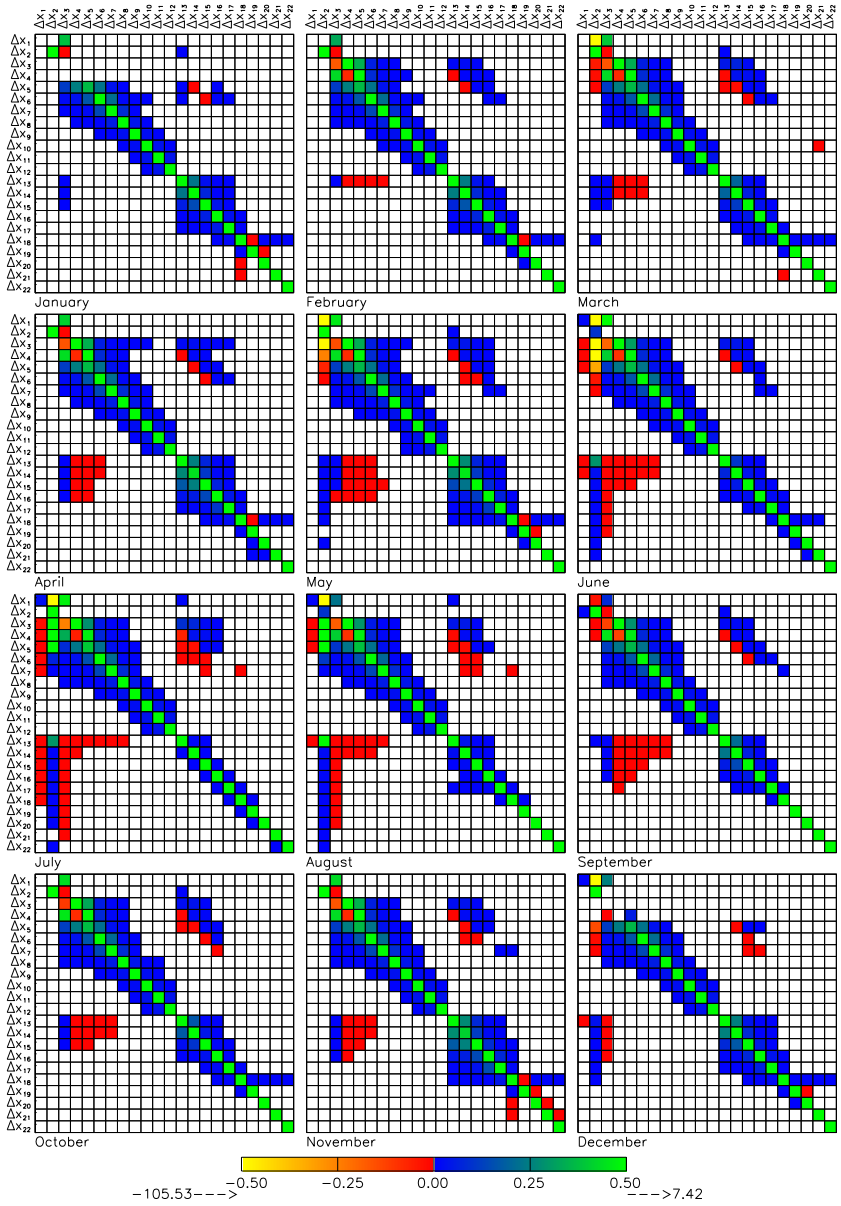


Figure 5.1: Monthly linearized system matrices during 2002 for CLM2.0 run for the OPE^3 field using default parameter values and expressions and OPE^3 meteorological forcings and surface data. x_1 is vegetation temperature, x_2 vegetation water, x_3 through x_{12} soil temperature and x_{13} through x_{22} soil moisture at 10 layers. A change in the state variable at time step i (horizontal axis), causes a change in the state variables at time step $i+1$ (vertical axis).

5.3.1 Input: land surface data, parameterization and initialization

By default, for initial state assignment and parameterization, several input files are used in CLM2.0 and some calculations are nested within the code:

- Land surface data (Surface-data.010x010.nc) need to be prepared at the desired grid and resolution. These data include: (i) geographical coordinates for all grid cells; (ii) soil color for all grid cells (by default, the same as in BATS); (iii) percentage clay and sand with vertical profile for all grid cells; (iv) land cover types and their respective fractions of the grid cell for all patches in all grid cells; (v) monthly averaged heights of the top and the bottom of the vegetation for all patches in all grid cells and (vi) monthly averaged LAI and SAI values for all patches in all grid cells.

  In this small scale study, the surface-data were always generated by the user and are summarized in table 5.1.

- As opposed to time-varying vegetation parameters defined in the surface datafile, time-invariant parameters for all plant functional types (pft) are stored in a simple ASCII file (pft-physiology). Each line in this file lists all optical, morphological and physiological parameters of one vegetation type. They are listed in table 5.2. By default, the land cover types are based on the International Geosphere-Biosphere Programme (IGBP) classification system.

- Time-invariant physical constants that do not depend on the soil texture and depth, called 'tunable parameters' in the CLM2.0 code (IniTimeConst.F90), are assigned default values that approximate physically realistic values. They are listed in table 5.3. The default value for the saturated soil hydraulic conductivity at the soil column bottom (***kd***) seems to be physically unrealistic, but in the manual it was reported to work well for CLM2.0 simulations.

- By default, soil physical parameters are calculated by CLM2.0 (IniTimeConst.F90) from soil texture and color, both defined in the surface data. The soil thermal and hydraulic parameters and soil albedos are listed in table 5.3.

- Besides this soil-vegetation parameterization, an initialization of the state variables (IniTimeVar.F90) is required. This can be done by using observed values. If such observations are not available, a spin-up run starting from arbitrary values should be executed. The state variables are also listed in table 5.2.

5.3.2 Input: atmospheric forcings

Atmospheric boundary conditions were stored with an hourly time interval in 1 file per month (e.g., 2001-01.nc for January 2001). In this work, the forcing data were

observed from 3 different meteorological towers and meteorological data are available from January 1, 2001 until today. The required data and the processing of these data were discussed in section 3.4. The observed atmospheric forcings were assumed to be uniform in space and kept identical for all patches during initialization and calibration. This erroneous assumption was made, even though it is well known that mainly the spatial variability of precipitation influences the model output and that lack of spatial information in the calibration procedure causes the optimal parameter sets to compensate for these errors in input. However, since the study area is smaller than the size of a typical precipitation cell [De Lannoy *et al.*, 2005], the spatial variability of rainfall will be limited.

5.4 CLM2.0 initialization and calibration for soil moisture

5.4.1 Global optimization method

In many optimization techniques, the shape of the objective function is sought and parameters are automatically tuned (iteratively) to find an absolute minimal value of the objective function. This requires an evaluation of the model output for every new combination of parameters. In this work a purely random, Monte Carlo (MC) sampling was performed to find the best parameters (171) and initial state variables (22) per profile. In total (171+22)*36=6948 parameters and initial states for all profiles were sought. Obviously, through the MC sampling it was only tried to get a rough idea of the parameters and initial states and a more advanced optimization method could be proposed. However, the direct advantage of simple MC sampling in the study at hand was that the methodology could be easily extended to serve for ensemble generation in next chapters. Furthermore, most existing advanced optimization algorithms were not designed to deal immediately with large amounts of parameters and it was beyond the scope of this research to adapt or extend these existing source codes.

Some drawbacks are that the method is rather inefficient and 'searches' (or better: 'samples') unwisely, disorganized and without memory through the generated MC samples to find the best one. Further, the number of needed MC simulations increases with the number of parameters. Even though it can be shown [Solomatine, 1995] that the pure direct random search offers an asymptotic guarantee to find the global optimum in the probabilistic sense, Duan *et al.* [1992] has shown that in practice for conceptual rainfall-runoff models, random search based global optimization procedures will fail to locate a global optimum with an acceptable probability of success. They found that minor local optima (at the detailed scale in the objective function) can only be detected by numerical analyses, while the random search will in general only find the optima at the overall scale.

The CLM2.0 was run forward for each patch or MC member with different 'param-

eter' vectors (a 'parameter' vector contains all model parameters and the initial state variables) and each corresponding time series of soil moisture was compared to observations through an objective or cost function. Each MC simulation represented a sample on the response surface of the likelihood measure within the model space and the samples of the response surface were projected onto a single parameter set axis. This allowed for a reduction of the multi-dimensional parameter space to a set of parameter combinations for which the value of the objective function was calculated (figure 5.2). The best parameter vector was simply the one with the lowest value for the objective function. The parameter vector contained all parameters listed in tables 5.2 and 5.3 for all modeled soil depths.

Traditional MC simulations take parameters from a uniform distribution. However, in this research, the parameters were drawn from Gaussian distributions, which were truncated, in order to not include physically or numerically impossible parameter values. A Gaussian distribution was chosen, as for most parameters the exact distribution representing the uncertainty on the parameters is not known, and it can be expected that the shape of the distributions is different for each parameter. Through some experiments using uniform distributions for the parameter perturbation it was found in this study that more resulting parameter combinations were invalid for use in the model, since there was a higher risk of combining (sometimes contrasting) extreme values for different parameters. Through Gaussian perturbation, a higher probability for an appropriate mean value results in a higher probability to obtain physically realistic parameters, or parameters that are effective for the model. By trial and error and studying numerical and physical problems in the model results, the standard deviation for perturbation was chosen to be maximally of the same order of magnitude as the mean value. Some boundary limits were imposed to avoid impossible values, such as, e.g., negative values for the hydraulic conductivity. When a parameter value beyond some predefined limiting bounds was generated, a new value was drawn.

The MC method was particular attractive in this work, as there is a possibility to run the CLM2.0 for different patches (instead of multiple restarting) on several processors, which reduces the computational time. Each patch represents a MC simulation. During this research a maximum of 4 processors was available for parallel computing, reducing the CPU with an approximate factor of 3.5 compared to a run on 1 processor. A rectangular grid of 10×10 cells was chosen over the area bounded by the outer boundaries of the OPE^3 field (northern coordinate of 39.03222 N, eastern coordinate of 76.84229 W, southern coordinate of 39.0252 N and western coordinate of 76.8475 W), with each grid cell containing 1500 patches. Through this combination of grid cells and patches, $15 \cdot 10^4$ MC simulations (= $10 \times 10 \times 1500$ patches) were generated, and from this collection of simulations, a best parameter set was extracted for each probe or point profile in space. The assignment of the parameters and initial state values is discusses below. The real geographical position of the grid cells and patches did not matter in this small scale study as all cells can be assigned identical features, and only a slight variability may be caused by a little differences in inclination angle of the sun, which were

negligible for this study. Further it should be recalled that the CLM2.0 does not represent lateral fluxes.

5.4.2 Uncertain initial conditions and parameters

Because the calibration of the model parameters is dependent on the choice of the initial state, an initial state estimation was included in the optimization procedure. Basically, this approach transformed the classical calibration problem into a weak-constraint variational data assimilation problem, in which both an optimal estimate of the initial state and of the parameters were sought. Through this state and parameter estimation it was hoped to obtain a model and parameters that would yield output with zero mean bias and purely random errors with minimum variance and no autocorrelation.

All parameters in the surface data file and in the pft-physiology file, the parameters calculated within the code and initial state values are uncertain numbers (except for the geographical coordinates and the fractions of the patches in the surface data file). Dai *et al.* [2003] mention in their overview of the input data for the CLM2.0 that the estimation of soil parameters is critically important and that users of CLM2.0 can assign different values than those a priori based on empirical relationships from earlier experience.

All parameters and initial state values were assumed to be Gaussian distributed around a highly likely value, based on measurements, information in literature or default values in the source code (a priori information). See tables 5.1, 5.2 and 5.3 for the mean values, standard deviations and possible boundary limits. The main idea of the distributions for the parameters over the grid cells and patches is illustrated in figure 5.3 and discussed below.

- In the surface data-file 10 grid cells in each direction (100 in total) were

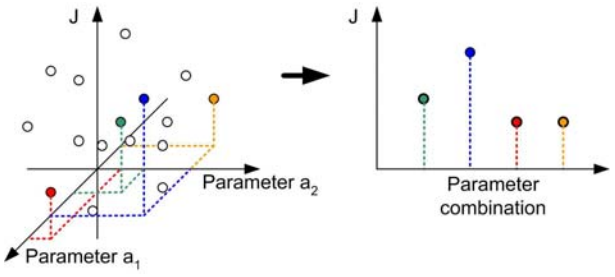


Figure 5.2: Monte Carlo optimization. Left: each combination of parameters in a multi-dimensional parameter space corresponds to a sample on the response surface. Right: each combination of parameters is projected to a single parameter set with a corresponding response value.

assigned Gaussian randomly distributed parameter values. So for each parameter/variable there were 100 realizations of soil texture (3 of the 100 are schematically shown in figure 5.3).

The mean and standard deviation for the upper (9 cm) soil texture in the OPE^3 field was calculated based on the data from the ARS-USDA. Profile information was obtained from point measurements at the SCAN-site of the NRCS-USDA (table 3.1). These data were presented in section 3.3.2.

For each patch, the monthly LAI was based on a combination of the default values and literature information [Boedhram *et al.*, 2001]. The monthly averaged height of the corn was calculated based on measurements taken during the intensive field campaign in 2002 as shown in figure 3.8 in section 3.3.3. For the last 2 variables, 1500 different realizations were generated at each grid cell, as 1500 patches were introduced in each grid cell.

- Each grid cell was subdivided into 1500 patches, representing 1500 different vegetation types. The vegetation types were identical for each individual grid cell. Plant physiological parameters are summarized in table 5.2. Information on these parameters for corn can be found in Hurtalová and Matejka [1999], Hurtalová *et al.* [2002], Fischenich [2000], Verhoef *et al.* [1997], Cox *et al.* [1998] and Earl and Tollenaar [1997]. It is clear that the sum of reflectance and transmittance for leaves and stem cannot exceed unity in any circumstance: here some dependency between parameters was allowed. It should be noted that corn assimilates carbon dioxide via the C4 photosynthetic pathway, which vegetation type is not included in the CLM2.0 datasets by default. Therefore, a new corn pft was introduced with the plant physiological parameters of corn as defined by default and adapted for the photosynthetic pathway only. Of course the photosynthetic pathway was kept invariable during calibration.

- For the time-invariant physical constants or tunable parameters that do not depend explicitly on (profile) texture by any equations, the default values of CLM2.0 were used as the mean and the standard deviation was chosen in the same order of magnitude. For all patches in all grid cells, a different realization for these constants was calculated.

- For all patches in all grid cells, time-invariant soil physical constants that vary over the soil profile were perturbed around a mean value obtained by calculating the empirical expression in the CLM2.0 code for that constant for a given texture in the grid cells. The mean values given in table 5.2 were calculated for 15.62 % clay and 62.17 % sand for reference. Information on these parameters can be found in Clapp and Hornberger [1978], Cosby *et al.* [1984] and Saxton *et al.* [1986].

- The initial state variables were perturbed around the mean that is proposed by default for spin-up runs in CLM2.0, except for the temperature, which was decreased from 283.0 K to 275.0 K as the run starts on January 1, 2001.

In this study, initial state values were included in the parameter estimation procedure, avoiding that optimal parameter estimates would compensate for misspecified initial conditions.

A limited sensitivity analysis was conducted to find the most important parameters for calibration. For all possible values (Gaussian distribution) of a given parameter, the corresponding values of an objective function for a simulation were plotted. However, not much of relevant information could be extracted from these plots for any of the objective criteria used. It is highly probable that this is due to complicated interactions between the parameters, which were all perturbed simultaneously and from which one parameter was selected per plot or because the parameter range was still chosen too small. Because of the difficulty to perform a decent selection of parameters to include in the calibration, it was decided to calibrate with brute force, perturbing all parameters without any effort to reduce the dimension of the parameter space. The advantage was that there was less risk to find improper parameters due to badly assigned constant values to some parameters. However, the problem is that adding additional parameters increases the minimum attainable uncertainty on the individual parameter estimates (cfr. Cramer-Rao ratio) [Cramér, 1946].

5.4.3 Calibration period

The model runs were initiated on January 1, 2001. A calibration period of 1 month was chosen in September 2001 (from September 2, 2001 through October 1, 2001). In this period, observations showed no evidence of lateral flow, as could be observed for some preceding months. Including this phenomenon would result in parameters that try to compensate for structural model errors, since the model does not simulate horizontal water flow. The state variables during the 24 hours on May 3, 2001, were chosen as the 'initial conditions', and their optimal values were obtained through smoothing by the variational assimilation (see below). The idea is schematically presented in figure 5.4. The remaining part of the observational data was used to study the model performance in predictive mode (validation).

The misfit between the modeled initial state during the day of May 3, 2001, and the observation was penalized two orders of magnitude more (factor 100) than the misfit during the month September, to mimic the common practice of using observations as best guess for the initial conditions. Furthermore, this helped to avoid that the best parameters would compensate for a badly initialized state vector. This method assured a balanced and realistic initial state at the start of the calibration period, without having to perform a spin-up of several years.

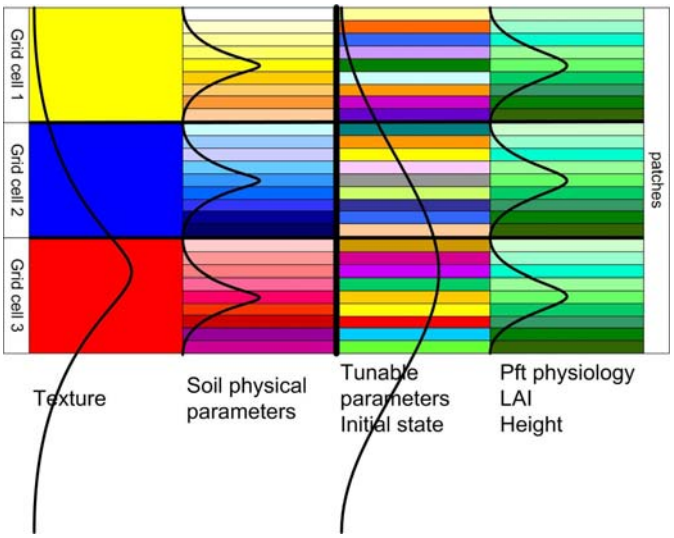


Figure 5.3: Gaussian perturbation of parameters in grid cells and patches. Column 1: each grid cell (100) is assigned 1 type of texture, which is constant for all patches inside the cell. Column 2: for each grid cell, the soil physical parameters are calculated based on this type of texture and this value is used as the mean around which perturbed parameters are calculated for each of the 1500 patches. Column 3: to each patch (independent of the grid cell) random tunable parameters and initial state variables are assigned. Column 4: in each grid cell the same 1500 vegetation types are present.

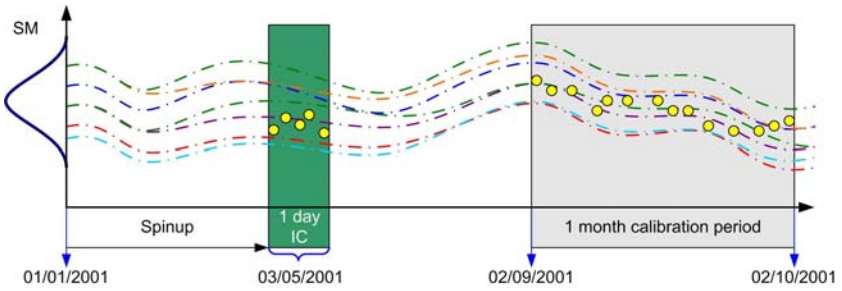


Figure 5.4: CLM2.0 initialization and calibration period. Several MC simulations are run forward in time and at each time step a distribution of soil moisture (SM) values is generated by the model. The best MC run will be extracted through comparison with observations (dots).

Table 5.1: Surface-data in CLM 2.0. Upper limits were not set, because no impossibly high parameter values were sampled.

	μ	σ	Min	Surface-data
PCT_SAND_1	62.17	5.56	0.	Percentage sand at 0.025 m [%]
PCT_CLAY_1	15.62	1.63	0.	Percentage clay at 0.025 m [%]
$SOIL\_COLOR$	-	-	-	1,2,3,4,5,6,7,8
PFT	-	-	-	17,18,...,1517
$MONTHLY\_LAI$	0.1	0.01	0.01	Monthly averaged LAI, Jan. [-]
$MONTHLY\_LAI$	0.1	0.01	0.01	"" LAI, Feb. [-]
$MONTHLY\_LAI$	0.1	0.01	0.01	"" LAI, Mar. [-]
$MONTHLY\_LAI$	0.2	0.01	0.01	"" LAI, Apr. [-]
$MONTHLY\_LAI$	0.5	0.1	0.01	"" LAI, May [-]
$MONTHLY\_LAI$	1.5	0.5	0.01	"" LAI, Jun. [-]
$MONTHLY\_LAI$	3.5	0.5	0.01	"" LAI, Jul. [-]
$MONTHLY\_LAI$	4	0.5	0.01	"" LAI, Aug. [-]
$MONTHLY\_LAI$	3.5	0.5	0.01	"" LAI, Sep. [-]
$MONTHLY\_LAI$	0.5	0.1	0.01	"" LAI, Oct. [-]
$MONTHLY\_LAI$	0.1	0.01	0.01	"" LAI, Nov. [-]
$MONTHLY\_LAI$	0.1	0.01	0.01	"" LAI, Dec. [-]
$MONTHLY\_HEIGHT\_TOP$	0.01	0	0.01	"" top height, Jan. [m]
$MONTHLY\_HEIGHT\_TOP$	0.01	0	0.01	"" top height, Feb. [m]
$MONTHLY\_HEIGHT\_TOP$	0.01	0	0.01	"" top height, Mar. [m]
$MONTHLY\_HEIGHT\_TOP$	0.01	0	0.01	"" top height, Apr. [m]
$MONTHLY\_HEIGHT\_TOP$	0.13	0.05	0.01	"" top height, May. [m]
$MONTHLY\_HEIGHT\_TOP$	0.85	0.16	0.01	"" top height, Jun. [m]
$MONTHLY\_HEIGHT\_TOP$	2.06	0.3	0.01	"" top height, Jul. [m]
$MONTHLY\_HEIGHT\_TOP$	2.20	0.3	0.01	"" top height, Aug. [m]
$MONTHLY\_HEIGHT\_TOP$	2.15	0.3	0.01	"" top height, Sep. [m]
$MONTHLY\_HEIGHT\_TOP$	0.01	0	0.01	"" top height, Oct. [m]
$MONTHLY\_HEIGHT\_TOP$	0.01	0	0.01	"" top height, Nov. [m]
$MONTHLY\_HEIGHT\_TOP$	0.01	0	0.01	"" top height, Dec. [m]

5.4.4 Objective functions

Different measures of goodness-of-fit can be defined and the optimal parameter combination will differ depending on the choice of the measure. Parameter combinations that produce good model results for multiple objective functions were sought. This formed a first aspect of the multi-objective calibration procedure applied here. The Root Mean Square Error ($RMSE$), Nash-Suttcliffe criterium (N_S), time series correlation coefficient (R), and absolute mean difference ($BIAS$) were chosen as objective functions and calculated over the 1 month calibration period in September for each MC simulation to quantify the correspondence in the temporal evolution of observed and simulated soil moisture. Note that in this work, the notation R for the correlation is reserved for the calibration and validation of model runs only. The coefficient of determination or the Nash-Sutcliffe coefficient [Nash and Sutcliffe, 1970] is commonly adopted for evaluating the goodness-of-fit of simulated hydrographs and is basically a transformed and normalized measure of the overall $RMSE$. Additionally, an RMSE with extra weights (factor 100) for misfits during the one day of initial conditions was

Table 5.2: Pft-related parameters and initial states in CLM 2.0.

	μ	σ	Min	Max	pft-physiology
$z0mr$	0.120	0.05	0	0.6	Ratio of momentum roughness length to canopy top height [-]
$displar$	0.68	0.05	0	1.0	Ratio of displacement height to canopy top height [-]
$dleaf$	0.04	0.01	0	-	Characteristic leaf dimension [m]
$c3psn$	0	0	-	-	Photosynthetic pathway: 0.=C4 (e.g. corn), 1.=C3 [-]
$vcmx25$	50	2	0	-	Max rate of carboxylation at 25°C [μ mol CO_2.m^{-2}.s^{-1}]
mp	9	0.5	0	-	Slope conductance-to-photosynthesis relationship
$qe25$	0.06	0.05	0	-	Quantum efficiency at 25°C [μ mol CO_2.μ mol photon^{-1}]
$rhol_{\lambda_v}$	0.11	0.05	0	(1-$taul_{\lambda_v}$)	Leaf reflectance: 1=vis [-] ($rhol(.,1)$)
$rhol_{\lambda_n}$	0.58	0.05	0	(1-$taul_{\lambda_n}$)	Leaf reflectance: 2=nir [-] ($rhol(.,2)$)
$rhos_{\lambda_v}$	0.36	0.05	0	(1-$taus_{\lambda_v}$)	Stem reflectance: 1=vis [-] ($rhos(.,1)$)
$rhos_{\lambda_n}$	0.58	0.05	0	(1-$taus_{\lambda_n}$)	Stem reflectance: 2=nir [-] ($rhos(.,2)$)
$taul_{\lambda_v}$	0.07	0.05	0	(1-$rhol_{\lambda_v}$)	Leaf transmittance: 1=vis [-] ($taul(.,1)$)
$taul_{\lambda_n}$	0.25	0.05	0	(1-$rhol_{\lambda_n}$)	Leaf transmittance: 2=nir [-] ($taul(.,2)$)
$taus_{\lambda_v}$	0.22	0.05	0	(1-$rhos_{\lambda_v}$)	Stem transmittance: 1=vis [-] ($taus(.,1)$)
$taus_{\lambda_n}$	0.38	0.05	0	(1-$rhos_{\lambda_n}$)	Stem transmittance: 2=nir [-] ($taus(.,2)$)
xl	-0.3	0.1	-1	1	Leaf/stem orientation index [-]
$roota\_par$	6.0	0.05	0	-	CLM rooting distribution parameter [m^{-1}]
$rootb\_par$	3.0	0.05	0	-	CLM rooting distribution parameter [m^{-1}]
	μ	σ	Min	Max	Initial state variables (start Jan. 01, 2001)
t_veg	275	5	273.16	-	Initial vegetation temperature [K]
t_soisno_j	275	5	273.16	-	Initial soil-snow temperature [K]
$h2ocan$	0	0.1	0	-	Initial water in canopy [-]
$h2osoi\_vol_j$	0.3	0.1	0.02	1	Initial soil moisture [-]
$h2osno$	0	0	-	-	Initial snow water [-]

Table 5.3: Parameters in CLM 2.0.

	μ	σ	Min	Max	Time invariant physical constants
zlnd	0.01	0.01	0.001	-	Roughness length for soil [m]
zsno	0.0024	0.001	0.0001	-	Roughness length for snow [m]
csoilc	0.004	0.001	0.0001	-	Drag coefficient for soil under canopy [-]
dewmx	0.1	0.1	0.0001	-	Maximum dew [mm]
wtfact	0.3	0.1	0	1	Fraction of model area with high water table [fraction]
capr	0.34	0.1	0.0001	-	Tuning factor to turn first layer T into surface T [-]
cnfac	0.5	0.01	0.001	1	Crank Nicholson factor between 0 and 1 [-]
ssi	0.033	0.01	0.001	1	Irreducible water saturation of snow [fraction]
wimp	0.05	0.01	0.005	1	Water impermeable if porosity $<$ ***wimp*** [fraction]
pondmx	10.0	5	0.001	-	Ponding depth [mm]
smpmax	-1.5×10^{5}	1.0×10^{5}	-	0.	Wilting point potential [mm]
smpmin	-1.0×10^{8}	1.0×10^{8}	-	0.	Restriction for min of soil poten. [mm]
trsmx0	2.0×10^{-4}	1.0×10^{-4}	0.	-	Max transpiration for moist soil+100% veg. [mm.s^{-1}]
fz	1.0	1	0	-	Water table depth scale parameter [m^{-1}]
kd	0.04	0.01	0	-	Saturated soil hydraulic conductivity bottom [mm.s^{-1}]
ld	1.0×10^{-5}	1.0×10^{-5}	0.	-	Base flow parameter for saturated fraction of watershed [mm.s^{-1}]
NwRb	5	1	int(10/3)	10-1	First bottom layer contributing to the calculation of base flow [-]
NwRs	3	1	1	***NwRb***	Last top layer contributing to the calculation of surface runoff [-]
hkdepth	0.5	0.1	0.0001	-	Length scale for Ksat decrease [m]
bsw$_j$	5.4	1.0	1	-	Clapp and Hornberger "b" [Clapp and Hornberger, 1978] [-]
watsat$_j$	0.41	0.1	0.005	1	Volumetric soil water at saturation (porosity) [-]
xksat$_j$	0.008	0.1	1.0×10^{-4}	-	Saturated hydraulic conductivity at surface [mm.s^{-1}]
hksat$_j$	0.0011	0.1	0	-	Hydraulic conductivity at saturation [mm.s^{-1}]
sucsat$_j$	116.30	50	0.5	-	Minimum soil suction [mm]
tkm$_j$	7.6	1	0.01	-	Thermal conductivity [W.m^{-1}.K^{-1}]
bd$_j$	1591	50	0.1	-	Bulk density [mg.cm^{-3}]
tkmg$_j$	3.3	1	1.0×10^{-4}	-	Thermal conductivity, soil minerals [W.m^{-1}.K^{-1}]
tksatu$_j$	2.6	1	1.0×10^{-4}	-	Thermal conductivity, saturated soil [W.m^{-1}.K^{-1}]
tkdry$_j$	0.23	0.1	1.0×10^{-4}	-	Thermal conductivity, dry soil [W.m^{-1}.K^{-1}]
csol$_j$	2179604	1.0×10^{6}	1	-	Heat capacity, soil solids [J.m^{-3}.K^{-1}]

calculated ($RMSE_{ic}$). These criteria or objective functions are given by:

$$RMSE = \sqrt{\frac{1}{\mathrm{T}} \sum_{i=b}^{b+\mathrm{T}-1} (y_i - \hat{x}_i)^2} \tag{5.2}$$

$$N\_S = 1 - \frac{\sum_{i=b}^{b+\mathrm{T}-1}(y_i - \hat{x}_i)^2}{\sum_{i=b}^{b+\mathrm{T}-1}(y_i - < y >)^2} \tag{5.3}$$

$$R = \frac{\sum_{i=b}^{b+\mathrm{T}-1}(y_i - < y >)(\hat{x}_i - < \hat{x} >)}{\sqrt{\sum_{i=b}^{b+\mathrm{T}-1}(y_i - < y >)^2 \sum_{i=b}^{b+\mathrm{T}-1}(\hat{x}_i - < \hat{x} >)^2}} \tag{5.4}$$

$$BIAS = \left| \frac{1}{\mathrm{T}} \sum_{i=b}^{b+\mathrm{T}-1} y_i - \frac{1}{\mathrm{T}} \sum_{i=b}^{b+\mathrm{T}-1} \hat{x}_i \right| = | < y > - < \hat{x} > | \tag{5.5}$$

$$RMSE_{ic} = \sqrt{\frac{1}{\mathrm{T} + \mathrm{Tic}} \left\{ 100 \sum_{j=a}^{a+\mathrm{Tic}-1} (y_j - \hat{x}_j)^2 + \sum_{i=b}^{b+\mathrm{T}-1} (y_i - \hat{x}_i)^2 \right\}} \tag{5.6}$$

with $\hat{x}_i$ the modeled state variables, y_i the corresponding observations and the subscript ic referring to the inclusion of initial conditions. The hourly time steps are denoted as $i \in [b, b + \mathrm{T} - 1]$ and $j \in [a, a + \mathrm{Tic} - 1]$, with $\mathrm{T} = 24 \cdot 30 = 720$ for the 1 month calibration period and $\mathrm{Tic} = 24$ for 1 day of initial conditions. Time step a corresponds to the first time step on September 2, 2001 and time step b corresponds to the first hour on May 3, 2001. The notation $<>$ refers to temporally averaged variables. Note that $\hat{x}_i$ could have been replaced by $h_i(\hat{x}_i)$, with h_i the function relating $\hat{x}_i$ to y_i, if the observations were no direct measurements of the modeled state. The expression for $RMSE_{ic}$ is similar to the least-square objective functions defined for variational assimilation, which makes that the proposed calibration with inclusion in initial state conditions is analogous to a weak constraint variational algorithm (with direct minimization of the cost or objective function).

For most probes in the field, different layers of observations serve for calibration of one soil column. This formed a second aspect in the multi-objective calibration. Measures of goodness-of-fit for the different layers were combined to one measure through aggregation, so that for each land column 1 value for a measure of goodness-of-fit was available. Aggregation through averaging of the values of the objective functions over all depths resulted in some compensation of the measures of goodness-of-fit for bad performing layers by the values of very well behaving layers. Therefore, it was studied whether it was possible to overcome this drawback by selecting the worst value of the objective functions over all depths and sort the best parameter sets based on this worst case criterium instead of on the averaged criterium. This caused that parameter sets for which all model soil layers performed very well and, for example, only one layer did perform badly for one criterion, were rejected in advantage of parameter sets which caused the model to simulate only moderately for all layers. Thus, selecting the worst criterium over

all layers in a profile did not result in an overall measure of goodness-of-fit over the whole profile. Therefore averaging of the values for the measures of goodness-of-fit over the different depths was preferred. To obtain the profile-average $RMSE$ and $RMSE_{ic}$, the individual layer $RMSE$s were weighed by the amount of pairs of observations and model outputs in each layer. So for each land column several profile-averaged measures were available.

In order to obtain a combined measure over all objective functions for each land column, aggregation of the different types of measures of goodness-of-fit was performed by computing the Euclidean distance, D, of the positions (d) that the MC simulations took relative to the best simulation after sorting (best to worst) on different criteria. For example, the third best MC simulation for $RMSE$ got a d_{RMSE} value of 3. The Euclidean distance is given by:

$$D = \sqrt{d_{RMSE}^2 + d_{RMSE_{ic}}^2 + d_{N\_S}^2 + d_R^2 + d_{BIAS}^2} \tag{5.7}$$

Another aspect in the multi-objectiveness of the calibration procedure would be the inclusion of different variables to calibrate for (e.g. soil temperature, evapotranspiration, runoff, ...). Except for the SCAN site, for which soil temperature data are recorded, only soil moisture observations are available in space over the OPE^3 field and thus the parameter and initial state optimization was limited to obtain the best estimates for soil moisture only.

5.4.5 Multi-objective calibration

Iterative sorting of the values for the different objective functions was performed and after each sort the worst patches were excluded for further competition. After a first sorting on the Euclidean distance D, half of the patches were excluded. As the selection proceeded for the individual measures of goodness-of-fit, less patches were excluded each time the sorting was performed. The sorting procedure passed twice through a series of selection criteria, with following numbers indicating the percentage (subjective choice) remaining patches after sorting on the following sequence of criteria: Euclidean (combined) distance D: 50 %, $RMSE_{ic}$: 50 %, N_S: 50 %, R: 50 %, $BIAS$: 70 %, $RMSE$: 70 %, $BIAS$: 50 %. Thus, starting from $15 \cdot 10^4$ patches, the last sorting algorithm (second pass through the sequence of criteria, ending with $BIAS$) worked on 35 patches only. Clearly, there was a subjective choice to limit the bias and a different sorting procedure would result in a different optimal parameter set, which refers to the equifinality concept [Beven, 1993; Beven and Freer, 2001], and the concept of multi-objective equivalence of Gupta *et al.* [1998].

This procedure is illustrated in figure 5.5 for probe BH1 and in figure 5.6 for probe CL2. The values of the measures illustrate that it was more difficult to calibrate for sensor CL2 with 7 observation depths than for sensor BH1 with only 3 observation depths. From this plot it is also clear that the number of considered

MC realizations was little less than 150000, since parameter vectors producing unrealistic soil moisture output (e.g. due to instability in the calculation scheme) were eliminated from further analysis. For the best 5 simulations, time series of soil moisture for probe BH1 are also shown in figure 5.5 and for some depths of probe CL2 in figure 5.6. Apparently, it was not possible to find a parameter set so that processes for the profile at probe CL2 were correctly represented by the CLM2.0.

The resulting averaged measures of goodness-of-fit are presented in table 5.10. Notice that these aggregated measures were calculated for different numbers of layers in a profile, depending on the available observations for each probe. The values reveal that calibration using only 3 depths of data (H-profiles) was usually easier than calibration for 6 or 7 depths. It was more difficult to find an optimal parameter set that yielded good model results for many soil layers than for only a few. Also, the deeper layers showed evidence of preferential flow, which cannot be captured by the model. However, the results for the model depths for which no observations were available for calibration, could not be checked, and may therefore deviate from the truth for the shallow profiles as well.

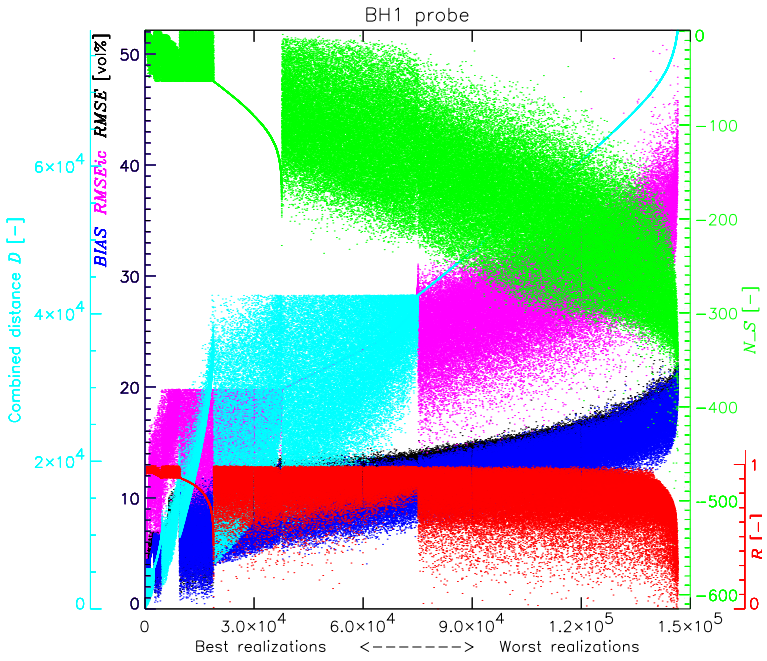
The proposed method was chosen instead of the approach of Houser *et al.* [2001], which also might have worked for this study, since the latter runs the risk of including parameter sets for which for example one measure may insinuate an extremely good performance, that cannot be found for any other parameter set, while for other measures rather low performances are found.

5.4.6 Multiple optimal parameter sets

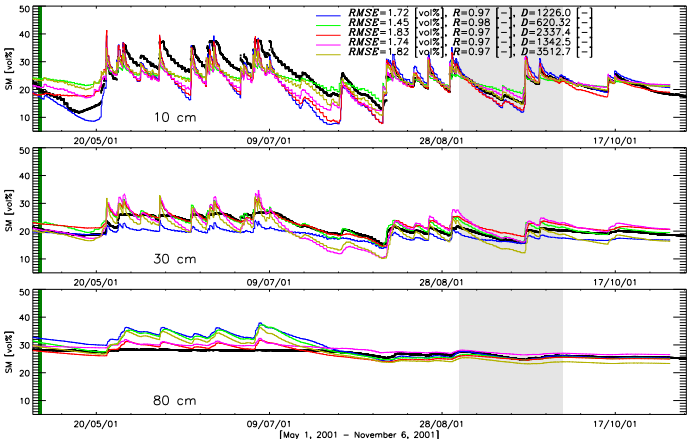
With the method above, a single best parameter set was obtained per probe. However, experimental tests revealed that changing the sorting procedure, resulted in other best parameter sets. In fact, many different parameter sets may result in good model simulations. In this context, there should be referred to section 4.3 and more in particular to the work of Gupta *et al.* [1998] who introduced the concept of multi-objective equivalence of parameter sets (Pareto optimal), and Beven and Binley [1992] who suggested that different parameter sets are likely to be equally probable, known to as the equifinality of parameter sets. An example of the 50 best parameter sets for the calibration of one profile is given in figure 5.7 and 5.8. For this figure all parameter values were rescaled as follows:

$$\text{normalized parameter value} = \frac{\text{optimal parameter value} - m \text{ parameter value}}{stdv \text{ of parameter value}} \tag{5.8}$$

with m and $stdv$ of the parameter value referring to the mean and standard deviation of the truncated Gaussian distributions for the perturbations of the parameters. From comparison of the best sets of optimal parameters it is clear that the actual values of the parameters span a broad range and that no single indicative value can be found. However, for most depths the range of best ***xksat*** and ***hksat***

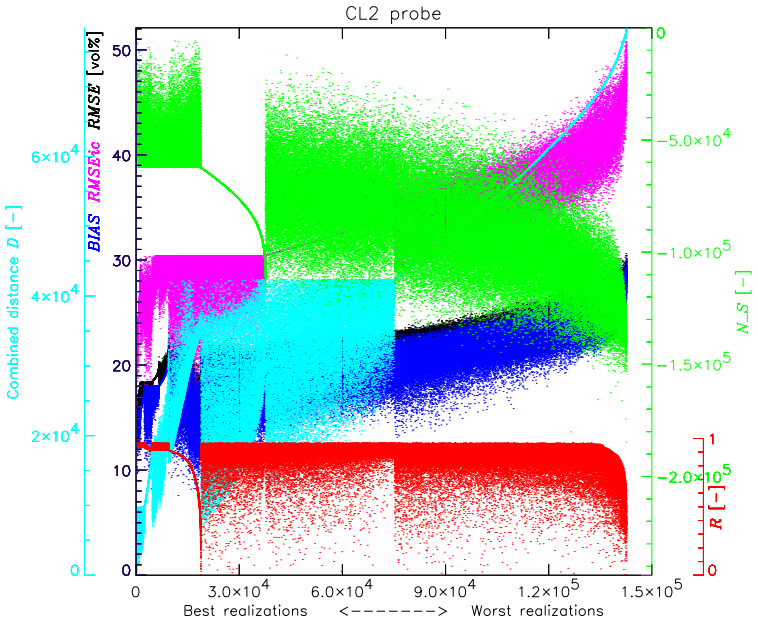


(a) Sorted MC simulations.

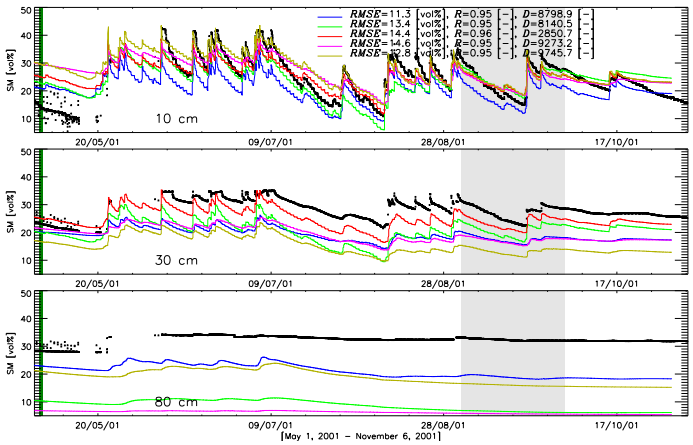


(b) Best 5 simulations.

Figure 5.5: Calibration of CLM2.0 for probe BH1. In the upper plot, the different profile-averaged measures and the combined distance measure D are sorted from best to worst. In the lower plot, the colored time series give the model results for the 5 best parameter sets, while the black dots are the observations. The green area indicated the period for which the initial conditions were estimated, the gray area is where the parameters were calibrated.



(a) Sorted MC simulations.



(b) Best 5 simulations.

Figure 5.6: Same as figure 5.5, but for probe CL2.

values was limited and also the initial soil temperature conditions were clearly preferred to be larger than the proposed mean.

Figure 5.9 compares how the MC simulations were sorted by different objectives and shows if there was any overlap in the best parameter sets for different criteria. Therefore, all simulations were sorted by each criterium separately. The sorted simulations were divided into classes of a given size S (which is represented by the different colors, 5 cases (30-150)). It was sought which fraction of the best S MC simulations according to one criterium were retrieved amongst the best S MC simulations (first class), the best S to $2S$ MC simulations (second class), the $2S$ to $3S$ MC simulations (third class), etc. according to another criterium. The x-axis shows the number of classes of S simulations that were checked. The y-axis shows the fraction of the S patches in the first class after sorting for one criterium which were retrieved in the different classes after sorting for a second criterium. Also the cumulated fraction is shown and this fraction reaches 1 after screening all classes of the second criterium. The fraction falling in the first class was in general not very high, except for the comparison of mean difference $BIAS$ and $RMSE$. However, in general, simulations yielding good values for one measure of goodness-of-fit, also resulted in good values for another objective function, except for the correlation R. So there was some agreement in the selection of the best patches for different criteria, in contrast to the several reports that one has not been able to find overlapping parameter sets when different objectives were used. However, it should also be recognized that the ultimate best patch was always different, depending on the method.

Table 5.4, 5.5, 5.6 and 5.7, show the mean, standard deviation, minimum and maximum values of each parameter over all 36 calibrated profiles. It is clear that a large spread on the parameters was found for the different soil profiles, even though they are all situated within a relatively small area.

The physical meaning of parameters could be discussed. The calibration period was chosen in a period of no clear preferential horizontal flow, which could not be captured by the model. However, the CLM2.0 does not model regular horizontal flow either, so parameters may have been found that try to compensate for example for this shortcoming. However, through the introduction of a priori knowledge in the sense of limits and Gaussian density functions for the perturbations of the parameters, the parameters had magnitudes that were physically realistic.

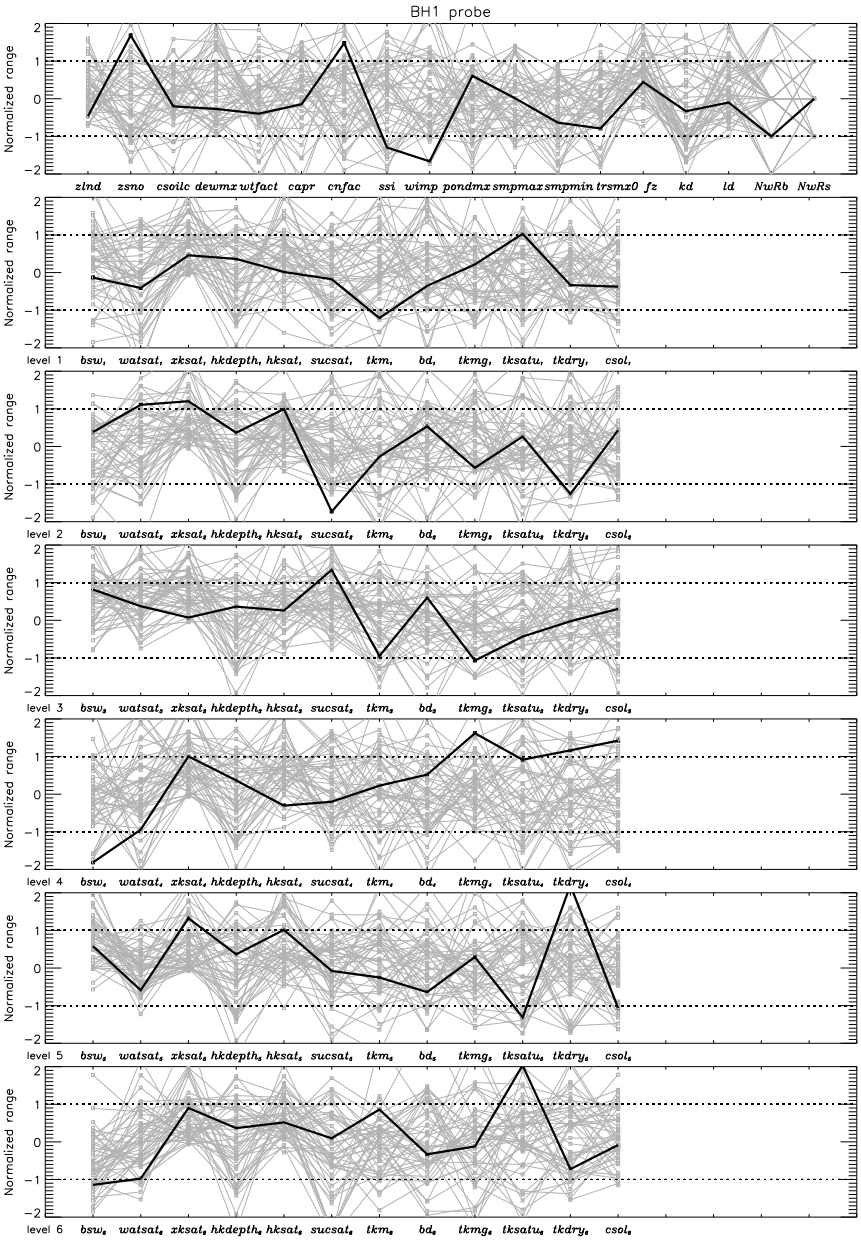


Figure 5.7: The best 50 parameter sets for probe BH1, with indication in thick black of the single optimal parameter set, which resulted from the sorting procedure. The horizontal dotted lines delineate the parameter range in between the mean - 1 standard deviation to the mean + 1 standard deviation. The levels refer to the modeled soil layers.

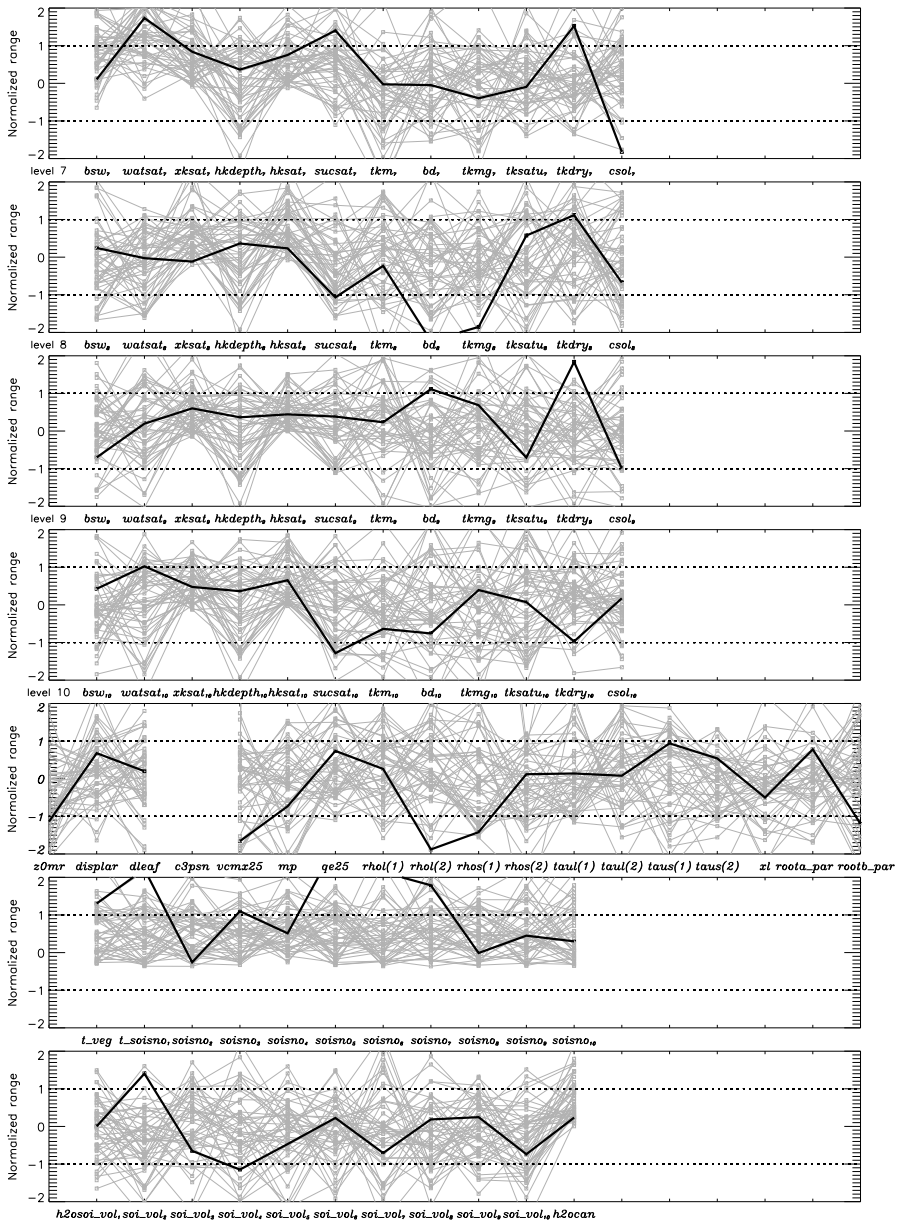


Figure 5.8: Figure 5.7 continued.

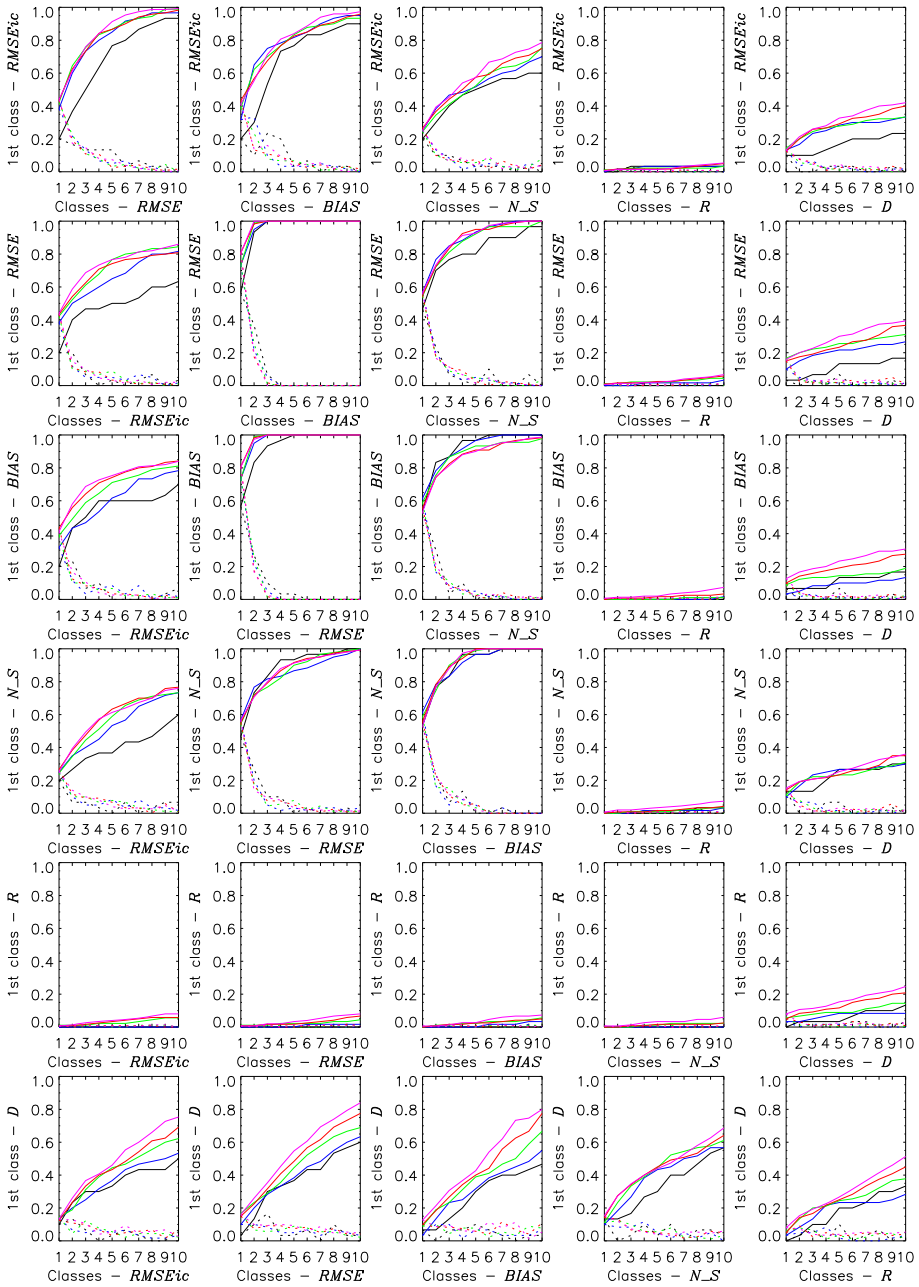


Figure 5.9: Comparison of sorting results by different objective functions for calibration and initialization at probe BH1, with a class size of 30 in black, 60 in blue, 90 in green, 120 in red and 150 in pink. Full lines are for cumulative frequencies, while dashed lines are for the frequencies in each class separately.

Table 5.4: Optimal surface parameters in CLM2.0 for the OPE[3] field, with m the mean and *stdv* the standard deviation over all profiles. The units correspond to those in table 5.1.

	m	*stdv*	Min	Max
SOIL_COLOR	4.611	2.348	1.000	8.000
MONTHLY_LAI, Jan.	0.102	0.010	0.078	0.122
MONTHLY_LAI, Feb.	0.103	0.011	0.081	0.124
MONTHLY_LAI, Mar.	0.098	0.010	0.078	0.121
MONTHLY_LAI, Apr.	0.197	0.009	0.175	0.219
MONTHLY_LAI, May	0.476	0.082	0.314	0.688
MONTHLY_LAI, Jun.	1.586	0.540	0.694	2.731
MONTHLY_LAI, Jul.	3.462	0.464	2.520	4.333
MONTHLY_LAI, Aug.	3.922	0.505	3.046	4.867
MONTHLY_LAI, Sep.	3.448	0.471	2.307	4.324
MONTHLY_LAI, Oct.	0.490	0.108	0.321	0.757
MONTHLY_LAI, Nov.	0.102	0.011	0.070	0.121
MONTHLY_LAI, Dec.	0.101	0.011	0.078	0.123
MONTHLY_HEIGHT_TOP, Jan.	0.010	-	0.010	0.010
MONTHLY_HEIGHT_TOP, Feb.	0.010	-	0.010	0.010
MONTHLY_HEIGHT_TOP, Mar.	0.010	-	0.010	0.010
MONTHLY_HEIGHT_TOP, Apr.	0.010	-	0.010	0.010
MONTHLY_HEIGHT_TOP, May	0.129	0.047	0.018	0.255
MONTHLY_HEIGHT_TOP, Jun.	0.847	0.121	0.604	1.109
MONTHLY_HEIGHT_TOP, Jul.	2.026	0.295	1.292	2.541
MONTHLY_HEIGHT_TOP, Aug.	2.187	0.320	1.420	2.805
MONTHLY_HEIGHT_TOP, Sep.	2.178	0.309	1.540	2.861
MONTHLY_HEIGHT_TOP, Oct.	0.010	-	0.010	0.010
MONTHLY_HEIGHT_TOP, Nov.	0.010	-	0.010	0.010
MONTHLY_HEIGHT_TOP, Dec.	0.010	-	0.010	0.010

Table 5.5: Optimal parameters in CLM2.0 for the OPE[3] field, with m the mean and *stdv* the standard deviation over all profiles. The units correspond to those in tables 5.2 and 5.3.

	m	*stdv*	Min	Max
z0mr	0.110	0.047	0.038	0.224
displar	0.696	0.048	0.585	0.778
dleaf	0.040	0.011	0.020	0.062
c3psn	0.000	0.000	0.000	0.000
vcmx25	49.858	2.303	46.169	53.946
mp	9.101	0.555	8.400	10.424
qe25	0.055	0.041	0.000	0.147
rhol(., 1)	0.096	0.047	0.006	0.203
rhol(., 2)	0.572	0.042	0.486	0.660
rhos(., 1)	0.369	0.044	0.289	0.475
rhos(., 2)	0.551	0.045	0.461	0.650
taul(., 1)	0.084	0.047	0.005	0.231
taul(., 2)	0.267	0.045	0.168	0.331
taus(., 1)	0.217	0.044	0.133	0.284
taus(., 2)	0.360	0.044	0.248	0.424
xl	-0.292	0.095	-0.466	-0.108
roota_par	6.005	0.049	5.892	6.097
rootb_par	2.995	0.043	2.905	3.105
zlnd	0.014	0.008	0.003	0.028
zsno	0.003	0.001	0.000	0.004
csoilc	0.004	0.001	0.003	0.006
dewmx	0.131	0.087	0.002	0.291
wtfact	0.264	0.082	0.097	0.463
capr	0.347	0.085	0.174	0.528
cnfac	0.497	0.009	0.471	0.515
ssi	0.032	0.010	0.015	0.056
wimp	0.048	0.012	0.023	0.067
pondmx	10.520	4.137	2.394	17.327
smpmax	-157.207×10^{3}	89.148×10^{3}	-403.212×10^{3}	-5.975×10^{3}
smpmin	-112.511×10^{6}	65.623×10^{6}	-294.587×10^{6}	-1.825×10^{6}
trsmx0	0.204×10^{-3}	0.079×10^{-3}	0.004×10^{-3}	0.307×10^{-3}
fz	1.340	0.773	0.065	2.854
kd	0.036	0.011	0.014	0.059
ld	0.013×10^{-3}	0.008×10^{-3}	0.000×10^{-3}	0.027×10^{-3}
NwRb	6.056	0.911	4.000	8.000
NwRs	3.222	0.946	2.000	5.000

Table 5.6: Table 5.5 continued. The units correspond to those in table 5.3. j is the modeled soil layer index.

	m	$stdv$	Min	Max
$\boldsymbol{hkdepth}$	0.483	0.100	0.252	0.658
j=1				
$\boldsymbol{bsw_j}$	3.375	0.746	1.633	5.128
$\boldsymbol{watsat_j}$	0.407	0.083	0.230	0.579
$\boldsymbol{xksat_j}$	0.083	0.067	0.001	0.277
$\boldsymbol{hksat_j}$	0.111	0.086	0.006	0.329
$\boldsymbol{sucsat_j}$	109.924	45.447	17.578	238.202
$\boldsymbol{tkm_j}$	8.919	1.034	6.297	10.819
$\boldsymbol{bd_j}$	1615.784	234.685	1138.067	2157.023
$\boldsymbol{tkmg_j}$	3.623	1.281	1.971	6.063
$\boldsymbol{tksatu_j}$	3.243	1.421	0.510	5.799
$\boldsymbol{tkdry_j}$	0.251	0.141	0.023	0.552
$\boldsymbol{csol_j}$	2135.265×10^3	866.870×10^3	358.465×10^3	4267.264×10^3
j=2				
$\boldsymbol{bsw_j}$	3.177	0.746	1.930	5.340
$\boldsymbol{watsat_j}$	0.406	0.096	0.145	0.551
$\boldsymbol{xksat_j}$	0.083	0.055	0.001	0.223
$\boldsymbol{hksat_j}$	0.153	0.110	0.016	0.388
$\boldsymbol{sucsat_j}$	94.929	56.384	4.400	205.866
$\boldsymbol{tkm_j}$	9.160	0.999	7.713	11.216
$\boldsymbol{bd_j}$	1611.654	255.822	1264.811	2356.198
$\boldsymbol{tkmg_j}$	4.018	1.257	2.186	6.825
$\boldsymbol{tksatu_j}$	3.548	1.478	1.436	7.223
$\boldsymbol{tkdry_j}$	0.242	0.157	0.038	0.797
$\boldsymbol{csol_j}$	2367.116×10^3	1048.696×10^3	622.617×10^3	4606.561×10^3
j=3				
$\boldsymbol{bsw_j}$	3.310	0.633	1.970	4.727
$\boldsymbol{watsat_j}$	0.451	0.081	0.321	0.604
$\boldsymbol{xksat_j}$	0.086	0.074	0.007	0.281
$\boldsymbol{hksat_j}$	0.111	0.066	0.003	0.242
$\boldsymbol{sucsat_j}$	147.786	42.738	54.228	237.341
$\boldsymbol{tkm_j}$	8.871	0.873	7.417	10.636
$\boldsymbol{bd_j}$	1475.200	234.805	1038.756	1868.623
$\boldsymbol{tkmg_j}$	3.329	1.003	1.186	5.943
$\boldsymbol{tksatu_j}$	2.594	1.164	0.093	4.687
$\boldsymbol{tkdry_j}$	0.218	0.076	0.017	0.427
$\boldsymbol{csol_j}$	2220.605×10^3	901.913×10^3	428.250×10^3	3614.716×10^3

Table 5.7: Table 5.6 continued.

	m	$stdv$	Min	Max
j=4				
$\boldsymbol{bsw_j}$	3.032	1.118	1.027	5.924
$\boldsymbol{watsat_j}$	0.437	0.097	0.252	0.612
$\boldsymbol{xksat_j}$	0.099	0.065	0.014	0.260
$\boldsymbol{hksat_j}$	0.105	0.076	0.003	0.281
$\boldsymbol{sucsat_j}$	119.341	31.089	47.292	176.225
$\boldsymbol{tkm_j}$	8.906	1.229	6.163	11.060
$\boldsymbol{bd_j}$	1505.564	264.753	1066.287	2044.085
$\boldsymbol{tkmg_j}$	3.760	1.033	1.561	6.038
$\boldsymbol{tksatu_j}$	2.901	1.409	0.423	5.946
$\boldsymbol{tkdry_j}$	0.230	0.146	0.009	0.657
$\boldsymbol{csol_j}$	2147.514×10^3	1168.845×10^3	227.341×10^3	5266.511×10^3
j=5				
$\boldsymbol{bsw_j}$	3.746	0.810	2.256	6.628
$\boldsymbol{watsat_j}$	0.475	0.069	0.330	0.601
$\boldsymbol{xksat_j}$	0.071	0.053	0.002	0.185
$\boldsymbol{hksat_j}$	0.068	0.052	0.001	0.185
$\boldsymbol{sucsat_j}$	153.753	33.245	101.031	202.545
$\boldsymbol{tkm_j}$	8.931	1.247	6.371	11.731
$\boldsymbol{bd_j}$	1412.135	189.630	1057.555	1772.265
$\boldsymbol{tkmg_j}$	2.792	0.867	1.131	4.783
$\boldsymbol{tksatu_j}$	2.243	1.050	0.260	5.262
$\boldsymbol{tkdry_j}$	0.233	0.113	0.036	0.483
$\boldsymbol{csol_j}$	2410.856×10^3	1082.396×10^3	656.651×10^3	5036.797×10^3
j=6				
$\boldsymbol{bsw_j}$	3.032	1.035	1.098	4.886
$\boldsymbol{watsat_j}$	0.473	0.094	0.297	0.637
$\boldsymbol{xksat_j}$	0.080	0.055	0.002	0.262
$\boldsymbol{hksat_j}$	0.096	0.079	0.001	0.279
$\boldsymbol{sucsat_j}$	131.064	56.671	38.625	246.486
$\boldsymbol{tkm_j}$	8.682	1.012	6.176	10.790
$\boldsymbol{bd_j}$	1421.499	259.251	972.677	1979.505
$\boldsymbol{tkmg_j}$	3.079	1.289	0.716	6.061
$\boldsymbol{tksatu_j}$	2.502	1.439	0.039	5.809
$\boldsymbol{tkdry_j}$	0.209	0.106	0.021	0.421
$\boldsymbol{csol_j}$	2262.266×10^3	870.176×10^3	502.337×10^3	4054.008×10^3
j=7				
$\boldsymbol{bsw_j}$	2.935	0.937	1.313	4.786
$\boldsymbol{watsat_j}$	0.466	0.118	0.219	0.616
$\boldsymbol{xksat_j}$	0.085	0.072	0.000	0.341
$\boldsymbol{hksat_j}$	0.080	0.065	0.002	0.159
$\boldsymbol{sucsat_j}$	110.031	49.678	8.113	220.096
$\boldsymbol{tkm_j}$	8.715	0.809	6.800	10.091
$\boldsymbol{bd_j}$	1449.552	320.354	997.275	2143.596
$\boldsymbol{tkmg_j}$	3.434	1.387	1.740	6.829
$\boldsymbol{tksatu_j}$	2.648	1.350	0.251	6.059
$\boldsymbol{tkdry_j}$	0.208	0.109	0.059	0.480
$\boldsymbol{csol_j}$	2230.235×10^3	1026.275×10^3	242.671×10^3	4043.718×10^3

Table 5.8: Table 5.7 continued.

	m	$stdv$	Min	Max
j=8				
$\boldsymbol{bsw_j}$	2.844	0.884	1.164	5.582
$\boldsymbol{watsat_j}$	0.419	0.096	0.257	0.608
$\boldsymbol{xksat_j}$	0.071	0.055	0.002	0.208
$\boldsymbol{hksat_j}$	0.092	0.084	0.000	0.265
$\boldsymbol{sucsat_j}$	81.109	44.734	11.009	194.579
$\boldsymbol{tkm_j}$	8.655	1.182	6.705	10.915
$\boldsymbol{bd_j}$	1546.745	263.958	1019.188	2049.114
$\boldsymbol{tkmg_j}$	3.573	1.119	1.907	5.955
$\boldsymbol{tksatu_j}$	3.134	1.246	0.919	5.579
$\boldsymbol{tkdry_j}$	0.234	0.122	0.034	0.530
$\boldsymbol{csol_j}$	2184.384×10^3	1091.135×10^3	408.389×10^3	4679.397×10^3
j=9				
$\boldsymbol{bsw_j}$	3.378	0.886	1.807	5.687
$\boldsymbol{watsat_j}$	0.429	0.096	0.209	0.616
$\boldsymbol{xksat_j}$	0.088	0.056	0.005	0.205
$\boldsymbol{hksat_j}$	0.059	0.056	0.006	0.152
$\boldsymbol{sucsat_j}$	100.772	42.899	12.540	199.915
$\boldsymbol{tkm_j}$	8.948	1.028	7.101	10.706
$\boldsymbol{bd_j}$	1534.627	263.958	1034.360	2178.024
$\boldsymbol{tkmg_j}$	3.741	1.374	1.576	7.062
$\boldsymbol{tksatu_j}$	3.382	1.651	0.509	6.787
$\boldsymbol{tkdry_j}$	0.249	0.131	0.050	0.621
$\boldsymbol{csol_j}$	1833.066×10^3	816.260×10^3	286.527×10^3	3455.346×10^3
j=10				
$\boldsymbol{bsw_j}$	3.345	0.775	2.085	4.853
$\boldsymbol{watsat_j}$	0.440	0.122	0.212	0.711
$\boldsymbol{xksat_j}$	0.095	0.058	0.002	0.214
$\boldsymbol{hksat_j}$	0.055	0.050	0.000	0.193
$\boldsymbol{sucsat_j}$	73.144	40.024	2.495	141.757
$\boldsymbol{tkm_j}$	9.012	0.907	6.464	10.324
$\boldsymbol{bd_j}$	1514.459	345.177	786.345	2186.431
$\boldsymbol{tkmg_j}$	3.364	1.124	0.559	5.552
$\boldsymbol{tksatu_j}$	3.024	1.577	0.105	6.095
$\boldsymbol{tkdry_j}$	0.250	0.120	0.020	0.583
$\boldsymbol{csol_j}$	1974.623×10^3	906.420×10^3	322.712×10^3	3812.251×10^3

Table 5.9: Table 5.8 continued. The units correspond to those in table 5.1.

	m	$stdv$	Min	Max
$\boldsymbol{t\_veg}$	277.498	3.049	273.618	283.634
$\boldsymbol{t\_soisno_1}$	278.309	3.578	273.196	286.388
$\boldsymbol{t\_soisno_2}$	278.329	3.672	273.711	288.711
$\boldsymbol{t\_soisno_3}$	277.605	3.260	273.245	285.217
$\boldsymbol{t\_soisno_4}$	277.655	2.711	273.256	283.211
$\boldsymbol{t\_soisno_5}$	277.642	3.603	273.204	289.187
$\boldsymbol{t\_soisno_6}$	279.085	4.028	273.701	290.673
$\boldsymbol{t\_soisno_7}$	278.351	3.511	273.342	285.463
$\boldsymbol{t\_soisno_8}$	277.334	3.110	273.217	284.230
$\boldsymbol{t\_soisno_9}$	277.940	3.604	273.297	288.501
$\boldsymbol{t\_soisno_{10}}$	277.722	3.781	273.186	292.065
$\boldsymbol{h2osoi\_vol_1}$	0.284	0.076	0.076	0.427
$\boldsymbol{h2osoi\_vol_2}$	0.282	0.088	0.031	0.435
$\boldsymbol{h2osoi\_vol_3}$	0.283	0.089	0.059	0.417
$\boldsymbol{h2osoi\_vol_4}$	0.280	0.087	0.129	0.449
$\boldsymbol{h2osoi\_vol_5}$	0.282	0.095	0.039	0.448
$\boldsymbol{h2osoi\_vol_6}$	0.326	0.089	0.081	0.515
$\boldsymbol{h2osoi\_vol_7}$	0.277	0.096	0.077	0.510
$\boldsymbol{h2osoi\_vol_8}$	0.253	0.078	0.085	0.406
$\boldsymbol{h2osoi\_vol_9}$	0.281	0.102	0.075	0.481
$\boldsymbol{h2osoi\_vol_{10}}$	0.286	0.066	0.117	0.457
$\boldsymbol{h2ocan}$	0.065	0.043	0.008	0.208

5.5 CLM2.0 validation for soil moisture

5.5.1 Validation period and validation measures

The validation period was chosen to start at October 3, 2001, after the month of September 2001, used for calibration. The complete remaining part (split sample) of the observed data set was used to calculate measures of goodness-of-fit for validation, i.e. through April 30, 2002. Because of the limited time period of data, it was not possible to include all possible dynamics of the watershed in either the initialization and calibration or the validation, and thus the calibration was performed during a period with growing vegetation and the validation was to be carried out during the non-growing season. For the validation the same (aggregated) measures of goodness-of-fit as for the calibration were used, except of course for the $RMSE_{ic}$ and the combined Euclidean distance (D), which included $RMSE_{ic}$ during calibration. Through the measures of goodness-of-fit as defined above, the temporal characteristics of observed and modeled time series of soil moisture were compared.

For further comparison to the observed time series of soil moisture, some spatio-temporal characteristics discussed in chapter 3 were determined for the modeled soil moisture. As opposed to the calculations for the validation of the different measures of goodness-of-fit, for these spatio-temporal characteristics, the entire year from May 1, 2001 through April 30, 2002 was included, which includes the initialization and calibration as well as the validation period. This allows direct comparison with the characteristics in chapter 3.

Time series of modeled soil moisture are shown in appendix B at the observation layers. Additionally, some time series of soil moisture at all modeled layers are shown in figures 9 through 16 in this appendix.

5.5.2 Results

Temporal characteristics

Measures of goodness-of-fit for validation
The values for the different measures of goodness-of-fit are summarized in table 5.10. Of course, also for validation, the values became worse when soil moisture at more depths was used for calibration. Measures of goodness-of-fit for the individual profile layers are plotted in figure 5.10. For probe CM2 only 2 instead of 6 measurement depths provided enough validation data and for probe AL2 only 6 of the 7 measurement depths were working properly. Per type of probe, the means of the relative values of the measures of goodness-of-fit for the individual layers were plotted in figure 5.11. For this figure, the values of the measures were linearly scaled between the maximum and minimum value for an individual layer

or sensor within the probe and these scaled values were averaged per depth over all probes of the same type (H-, M- or L-probes). In general, the deeper layers performed little worse than the other layers. The entire validation runs with optimal parameters for the different soil profiles will be referred to as the control runs in the remainder of this dissertation.

Note that the N_S values suggested overall bad performances and a better performance for the validation period than for the calibration period. The explanation is that for soil moisture in deeper layers, the temporal variability in the observations during the one month of calibration was minimal and hence it caused very large negative values of N_S for these layers. The averaged N_S over the whole profile (i.e. for the different layers of observations) was highly influenced by these extreme negative values. The validation spanned a longer time period and hence for the deeper layers the variability in soil moisture observations was larger and consequently the N_S for these deeper layers was less negative.

Plots of the resulting soil moisture profile evolution in time for all sensors are given in the appendix B. The plots are identically scaled to similar plots of the observations presented in appendix A. For these plots, the modeled soil moisture values were interpolated through cubic splines in between the observation depths to allow fair comparison with the plots for the observations. However, since the model calculates soil moisture at 10 depths for each sensor, more complete profiles of modeled soil moisture at each sensor can be plotted. These plots are also shown in appendix B. It seems that the model provided realistic soil moisture profiles at H-probes that were calibrated for 3 depths only.

Comparison of the residuals $(\hat{x}_i - y_i)$ between the modeled soil moisture values $\hat{x}_i$ and the observations y_i to the magnitude of the observed soil moisture (data not shown) revealed that the residuals varied with the magnitude of the observed soil moisture, most clearly for the upper soil layer and consequently, the residual was dependent on the soil moisture value. For almost all probes the residual for the upper layer soil moisture decreased with increasing soil moisture, with positive values for low soil moisture and negative values for high soil moisture values. For the upper soil layer, the model underestimated high soil moisture values and overestimated low soil moisture. On the other hand, the residuals were quite homoscedastic: the variance remained similar independent of the soil moisture value, which supports the use of the $RMSE$ as a measure of goodness-of-fit.

Autocorrelation and cross-correlation

The characteristic time scale of the modeled time series of soil moisture was for most sensors similar to the observed values. Bad correspondence in autocorrelation functions was found for probes AM1, BL1, BL2, BL3 and DL1. Typically, the autocorrelation functions for these sensors did not reach a stable (minimal) value within 3 months (for the upper layers), while for observed values the autocorrelation functions reached a stable (minimal) value after 50 days. Bad correspondence was also found for sensors CM1 and DH4, for which the autocorrelation functions for the modeled upper layer soil moisture reached a stable value after a month,

Table 5.10: Calibration and validation measures of goodness-of-fit.

	Calibration						Validation			
Probe	$RMSE_{ic}$ [vol%]	$RMSE$ [vol%]	$BIAS$ [vol%]	N_S [-]	R [-]	D [-]	$RMSE$ [vol%]	$BIAS$ [vol%]	N_S [-]	R [-]
AH1	7.96	1.10	0.65	0.64	0.95	3169.33	2.80	2.12	-3.03	0.78
AH2	4.49	1.75	1.07	-1.89	0.96	6074.03	-	-	-	-
AL1	22.14	9.53	6.55	-231.16	0.95	3835.89	7.22	4.73	-85.54	0.81
AL2	10.62	4.36	3.88	-836.11	0.87	3814.29	5.98	4.49	-56.86	0.33
AM1	15.35	7.00	5.28	-1184.29	0.97	6293.57	7.22	6.27	-369.76	0.67
AM2	9.82	5.16	3.72	-1036.07	0.98	2647.37	4.93	2.72	-104.79	0.83
BH1	5.73	1.73	0.68	0.40	0.98	1226.07	2.91	2.05	-2.37	0.87
BH2	6.92	1.55	0.50	0.18	0.98	3062.42	2.91	1.58	-0.76	0.87
BH3	5.84	1.38	1.13	-0.71	0.95	4959.59	2.16	1.28	-0.35	0.89
BH4	9.11	1.36	0.92	-0.63	0.96	3390.33	3.21	1.56	-0.14	0.86
BL1	20.40	4.88	3.74	-11867.83	0.77	6288.15	9.03	6.65	-117.03	0.47
BL2	19.56	4.46	3.61	-2158.60	0.79	2373.84	8.81	5.07	-6.13	0.55
BL3	17.05	4.21	2.81	-1028.43	0.95	6027.98	6.60	4.94	-232.98	0.69
BL4	12.87	5.57	4.53	-1775.54	0.95	4802.80	6.06	4.37	-132.96	0.89
BM1	7.24	3.94	2.72	-5838.98	0.79	5061.42	4.46	3.21	-71.67	0.86
BM2	18.76	7.95	6.15	-56364.58	0.47	1118.64	9.01	7.36	-1450.91	0.55
BM3	9.60	5.10	3.20	-4509.13	0.96	13035.31	4.75	3.15	-75.52	0.63
BM4	17.29	6.74	5.08	-154.29	0.97	1750.56	6.55	4.77	-190.74	0.87
CH1	10.56	3.13	1.25	0.35	0.97	3696.77	4.26	2.62	-0.30	0.80
CH2	8.48	1.79	0.87	-2.80	0.95	1215.18	4.68	3.76	-11.20	0.82
CL1	15.64	2.83	2.36	-50.36	0.94	8915.62	5.02	4.10	-9.42	0.78
CL2	15.26	11.38	10.34	-17761.21	0.95	8798.92	10.45	8.96	-273.11	0.84
CM1	31.40	15.17	12.00	-17759.44	0.57	2054.99	14.54	12.11	-570.11	0.66
CM2	10.83	2.24	0.53	-0.24	0.98	2511.38	4.35	2.86	-6.33	0.90
DH1	8.30	1.76	0.89	-4.53	0.95	2002.92	2.41	1.40	-1.11	0.87
DH2	6.76	2.17	1.14	0.00	0.98	2440.26	2.42	1.54	-1.26	0.88
DH3	16.01	2.20	0.89	-66.79	0.60	5459.91	2.62	1.48	-4.11	0.78
DH4	16.57	2.54	1.46	-32.25	0.95	7116.66	4.36	3.09	-7.15	0.84
DL1	7.07	3.58	3.07	-47.26	0.95	3938.01	2.86	2.05	-4.92	0.87
DL2	9.99	4.21	2.87	-202.82	0.73	722.63	5.16	3.33	-48.87	0.68
DL3	20.40	9.88	7.04	-3765.44	0.75	1968.55	9.86	8.33	-140.18	0.54
DL4	18.48	9.77	7.15	-20793.64	0.90	1530.06	9.29	5.99	-2396.01	0.74
DM1	6.07	3.83	2.72	-471.72	0.98	6707.11	3.84	2.63	-9.99	0.92
DM2	11.50	5.32	3.49	-563.11	0.97	8727.97	5.25	3.61	-34.29	0.42
DM3	9.21	4.76	3.76	-307.22	0.95	3171.75	5.19	3.69	-39.66	0.90
DM4	15.27	6.85	5.49	-3131.45	0.78	1498.22	8.28	6.99	-307.49	0.59

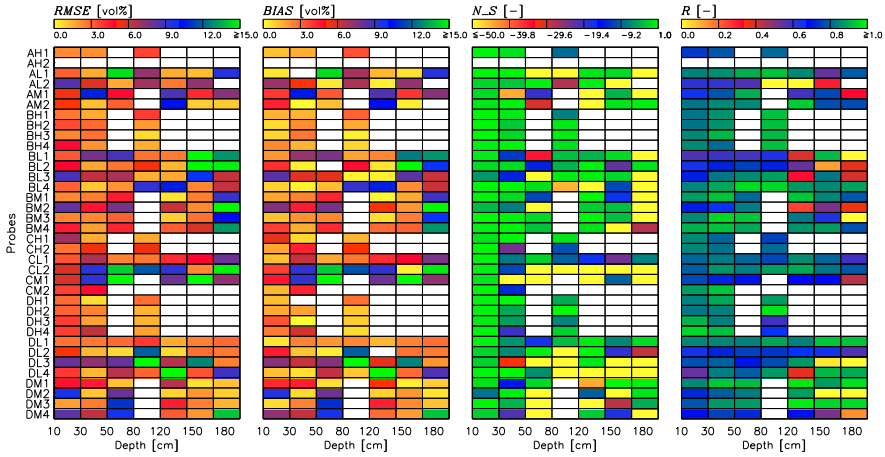


Figure 5.10: Validation measures of goodness-of-fit for the individual layers of all probes. White is where no or not enough validating observations are available.

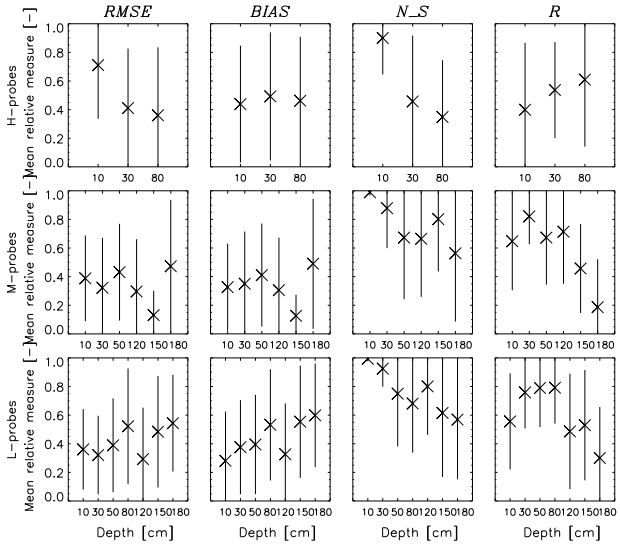


Figure 5.11: Mean relative values of different measures of goodness-of-fit for the different depths of the different types of probes.

while for the observed soil moisture, a stable value was only found after 100 days. In figure 5.12, autocorrelation functions for probes AM1 and BL4 are shown (compare to figure 3.17). The correlation lengths found for the modeled time series are summarized in table 5.11. In general these temporal correlation lengths were similar to the observed ones.

Table 5.11: Median M, minimum min, maximum max, mean m and standard deviation $stdv$ [days] of the temporal autocorrelation length, determined for all land columns (N) for which observing sensors are also available.

Depth	10 cm	30 cm	50 cm	80 cm	120 cm	150 cm	180 cm
M	17	20	44	49	53	57	62
min	3	4	13	17	14	17	18
max	51	52	56	81	105	111	111
m	19	26	35	44	56	60	63
$stdv$	14	14	14	14	23	26	26
N	36	36	24	24	24	23	21

In figure 5.13, (compare to figure 3.21) the absolute values of the difference between the correlation matrix for the observed and modeled time series are plotted. The model did not represent the fact that some sensors were really badly correlated to most other sensors, which caused a large difference in correlation. The observed 'special' behavior of some sensors was difficult to capture in a model, especially since the model only represents relatively simple vertical flow through a soil column. In figure 5.14, (compare to figure 3.25) the different sensors were ranked by their spatially averaged correlation. The correlation between the modeled time series differed a lot from the correlation between the observed time series. Apparently, the modeled soil moisture at probes AL2 and DL2 were badly correlated to the modeled soil moisture for other sensors for all depths, which refers to a problem in calibration, because the observed values for these sensors did not suggest a low correlation (definitely not for AL2).

Spatial characteristics

The spatially averaged soil moisture, the standard deviation and coefficient of variation are plotted in figure 5.15 (compare to figure 3.30). The standard deviation increased with increasing modeled soil moisture for all layers, which was not the case for the observed soil moisture. Less variability in time was found for the coefficient of variation obtained for the modeled time series than for the observed series. In table 5.12, the temporal averages of these spatial averages and standard deviations are summarized. Time averages of spatial statistics were taken over the time period of 1 year, winter time and summer time (compare to table 3.8). In general, the modeled time-space averaged soil moisture varied similar to the observed time-space averaged soil moisture over the profile, but for most depths (for all fields) the model tended to slightly underestimate soil moisture. Also the time averaged standard deviations in space compared well to the statistics found for the observations.

Table 5.12: Overview of the time-space averaged soil moisture [vol%] and the temporally averaged spatial standard deviation [vol%] for different periods for all sensors in the individual sub-watersheds, as well as in the complete OPE^3 field (All) and for the spatial averages of the 4 sub-watersheds within the complete field (4sub).

Depth [cm]	A *m*	*stdv*	B *m*	*stdv*	C *m*	*stdv*	D *m*	*stdv*	All *m*	*stdv*	4sub *m*	*stdv*
1 year (May 1, 2001 - April 30, 2002)												
10	22.9	5.6	19.4	4.0	22.3	2.4	22.5	4.8	21.5	4.7	21.79	1.4
30	25.5	5.4	23.1	4.3	29.6	6.4	27.0	4.1	25.9	5.4	26.3	2.4
50	30.5	6.4	21.8	4.8	17.7	5.5	26.8	4.1	24.3	6.6	24.22	4.9
80	22.5	4.9	19.1	8.7	15.8	9.5	21.6	6.6	19.9	8.0	19.73	2.7
120	14.6	2.8	14.4	8.4	16.8	5.0	16.4	7.6	15.5	7.0	15.54	1.1
150	17.6	6.0	24.5	9.2	17.2	11.6	18.2	10.0	20.0	10.0	19.37	3.0
180	19.5	7.2	20.5	4.3	21.9	7.1	18.7	8.2	19.9	6.9	20.14	1.2
summer period only (May 1, 2001 - October 30, 2001)												
10	22.8	5.5	18.8	4.0	21.8	2.5	21.8	5.0	21.0	4.8	21.28	1.5
30	25.7	5.3	23.0	4.2	29.1	6.1	26.7	4.2	25.7	5.3	26.14	2.2
50	31.6	6.6	22.4	4.9	18.1	5.5	27.0	4.2	24.8	6.8	24.77	5.1
80	23.5	4.9	20.0	8.8	16.6	10.0	22.6	6.5	20.9	8.1	20.68	2.7
120	15.8	2.8	15.5	8.0	18.0	4.8	18.1	8.3	16.8	7.1	16.86	1.2
150	19.1	6.0	26.6	10.0	18.3	12.0	19.4	10.3	21.6	10.5	20.84	3.4
180	21.4	7.6	22.8	4.3	23.6	7.5	20.2	9.1	21.9	7.4	22.02	1.3
winter period only (October 31, 2001 - April 30, 2002)												
10	23.6	6.2	20.5	4.1	23.1	2.3	23.3	4.8	22.4	4.7	22.65	1.3
30	25.9	5.5	23.9	4.4	30.5	6.4	27.8	4.2	26.6	5.5	27.03	2.4
50	30.7	6.0	22.6	4.8	18.0	5.5	27.3	4.1	24.7	6.5	24.62	4.8
80	22.2	4.6	19.4	8.7	15.8	9.6	21.9	6.6	20.1	8.0	19.81	2.6
120	14.1	2.9	13.9	8.6	16.5	5.0	15.8	7.3	15.0	7.0	15.06	1.1
150	16.8	5.9	23.5	9.0	16.8	11.6	17.6	9.9	19.3	9.8	18.68	2.8
180	18.4	7.0	19.2	4.4	21.0	6.9	17.8	7.8	18.9	6.7	19.1	1.2

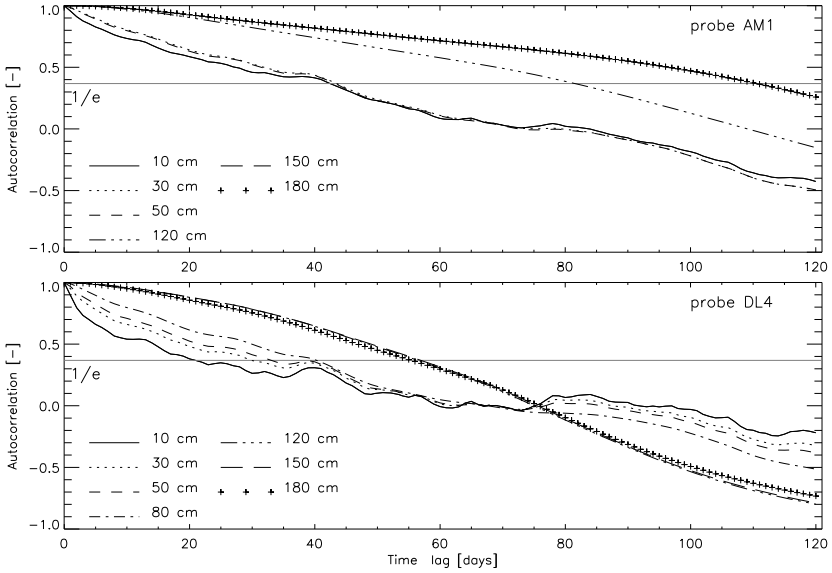


Figure 5.12: Autocorrelation for modeled soil moisture at probes AM1 and BL4 at all depths.

As in chapter 3, the spatial structure of modeled soil moisture was investigated by calculation of variograms for each time step, as well as after pooling for each month and for the period of one year form May 1, 2001 through April 30, 2002. In figure 5.16, the evolution of the range and sill for the modeled soil moisture are presented. Just as for the observed data, there were more difficulties to fit variogram models for the deeper layers. For the modeled soil moisture, also variogram fitting for the upper soil layer at 10 cm was often impossible. The values of the range and sill for the modeled and observed soil moisture were similar in magnitude. At 50 cm depth and occasionally at 30 cm depth, there was a decrease in range for increasing wetness. For the other layers, the evolution of range was proportionally with soil moisture or could not be discussed as variogram fitting was not possible. The sill mostly evolved also proportionally to soil moisture, except for 50 cm depth, where some increases in soil moisture resulted in a decreasing sill. For the parameters describing the monthly pooled variograms, some agreement with the observed variograms could be found at 30 cm (good), 50 cm (poor) and 80 cm (very good, see figure 5.17 and compare to figure 3.34). For the other soil layers, variogram fitting was not possible or (at 120 cm) resulted in very different parameter estimates than for the observed soil moisture. Mostly, the variances for small lag distances were not smaller than those for larger lag distances, which made the fitting difficult. However, even when no fitting was possible, for most depths the observed semivariances were situated within the range of calculated semivariances for the model results. Furthermore, when large/small variability in semivariances

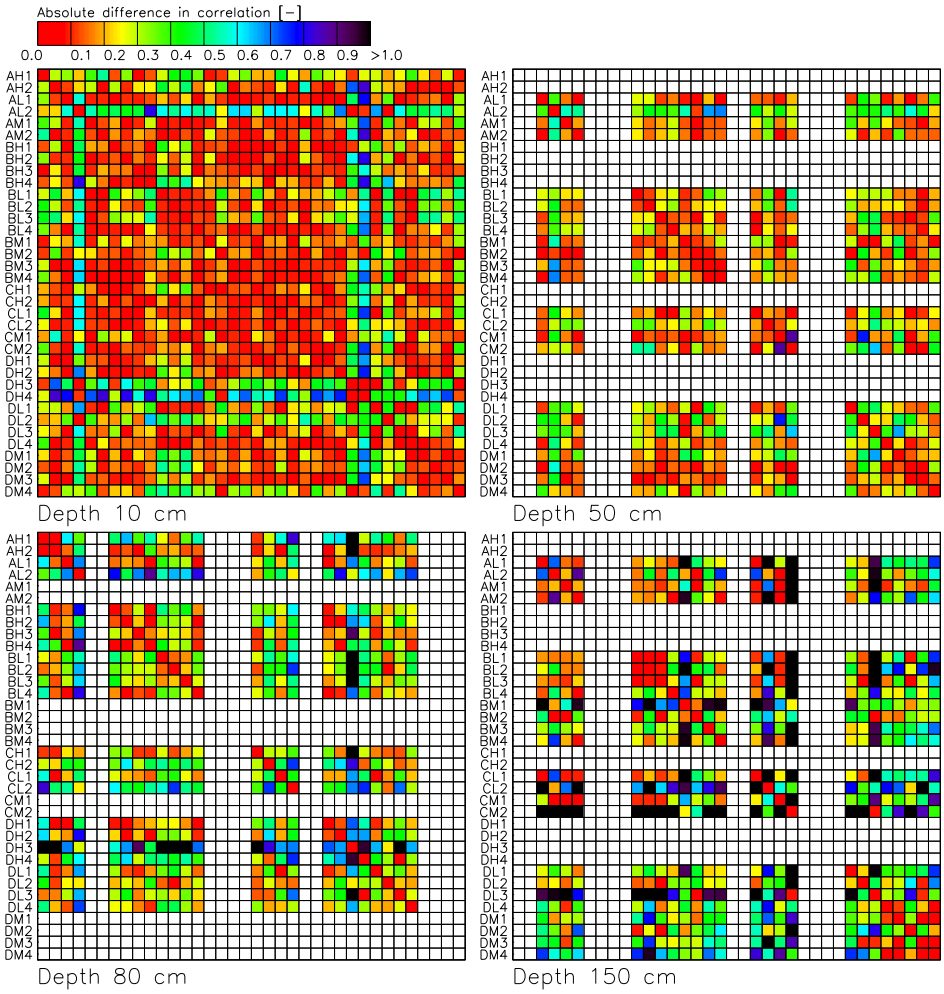


Figure 5.13: Absolute difference of correlation matrices at some observation depths of all sensors for the modeled time series during 1 year from May, 2001 to May, 2002. White is for missing observing sensors.

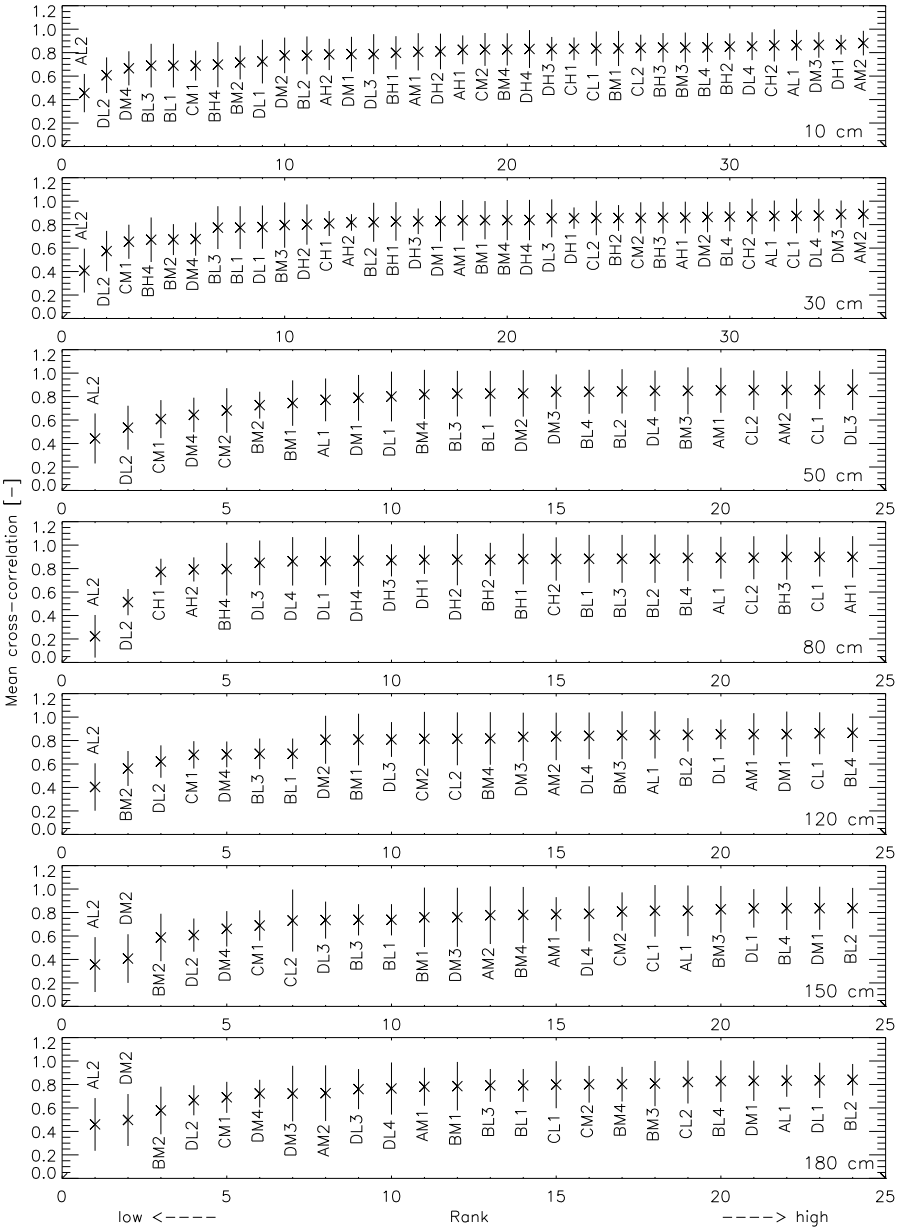


Figure 5.14: Ranked averaged cross-correlation and standard deviation for all sensors in the OPE^3 field.

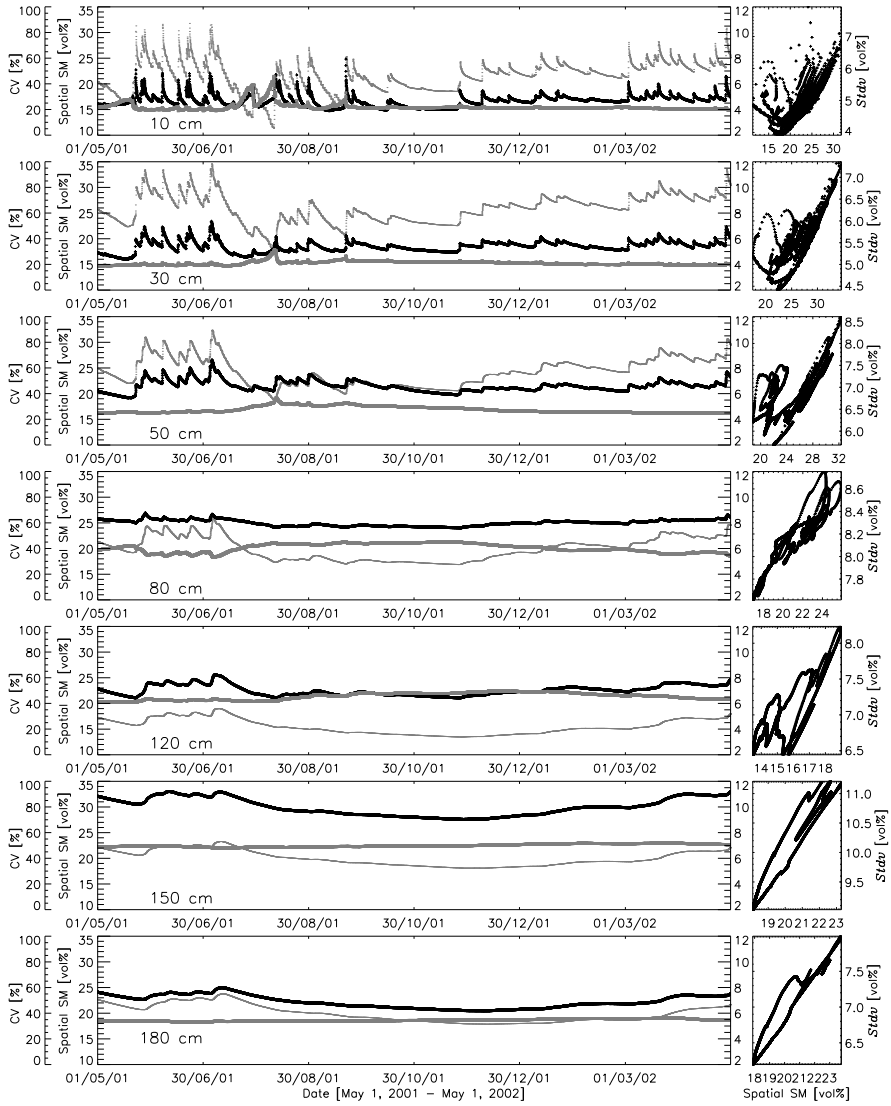


Figure 5.15: Spatially averaged modeled soil moisture SM (thin gray, left vertical scale), standard deviation *stdv* (black +, right vertical scale) and coefficient of variation CV (gray □, extra left vertical scale) at different depths for all probes in the OPE^3 field during the period May 1, 2001 through April 30, 2002. The righthand side figures plot the standard deviation, observed at the different time steps, in function of the spatial average soil moisture at the corresponding time steps.

was observed for the measurements within a month, also a large/small variability in semivariances for the model results was found during the same period. In table 5.13, geostatistical parameter estimates for pooled stratified (in time) variograms over 1 year are summarized. The sill values as well as the range values were lower than the sills and ranges estimated for observed soil moisture. Below 30 cm, the range decreased, similar as for observed soil moisture. The best agreement with the parameters for observed soil moisture was found at 80 cm depth.

Next, the time stability of the modeled sensors and their mean value relative to the spatial mean was studied (see figure 5.18 and compare to figure 3.36). For observed soil moisture, probes BL4 and AM1 were found to be representative for the evolution of the spatially averaged soil moisture. This was true for modeled soil moisture also for probe BL4, but for probe AM1 this was not the case. Examples of sensors that were really not representative for the observed spatial soil moisture, were probes DH4, BL1 and BL2. This was not valid for modeled soil moisture: soil moisture at these sensors apparently showed a more similar temporal evolution to the spatially averaged modeled soil moisture. In general, there was hardly any relationship between the standard deviation of difference ($stdv(d_j)$) for the observed soil moisture and the modeled soil moisture. At the other hand, sensors that were characterized by higher soil moisture values than the spatial average were also found to be characterized by higher modeled soil moisture values than the spatially averaged soil moisture. The individual watersheds were again very close to the complete field average soil moisture.

5.5.3 Model hydrology

The model was able to reproduce the general space-time characteristics of soil moisture over the complete OPE^3 field. However, 'anomalous', extreme or special profile-specific characteristics were hardly simulated. The time series dynamics for simulated profiles differed in general less amongst each other than the observed time series.

In the upper layers, the model performed very well, but the general tendency was towards an underestimation of high soil moisture and an overestimation of low soil moisture. The relatively good quality of the meteorological forcings guaranteed a realistic temporal evolution of soil moisture and served as the main controlling input to the modeled system. However, it was not possible to retrieve the ob-

Table 5.13: Geostatistical parameters describing the spatial structure of modeled soil moisture over the whole OPE^3 field. The range and sill are derived from fitting an exponential model to the pooled (over 1 year) stratified (in time) variograms calculated at lag 20 m.

lag 20 m	10 cm	30 cm	50 cm	80 cm	120 cm	150 cm	180 cm
Sill [vol%]	/	40	59	92	50	/	/
Range [m]	/	269	267	211	7	/	/

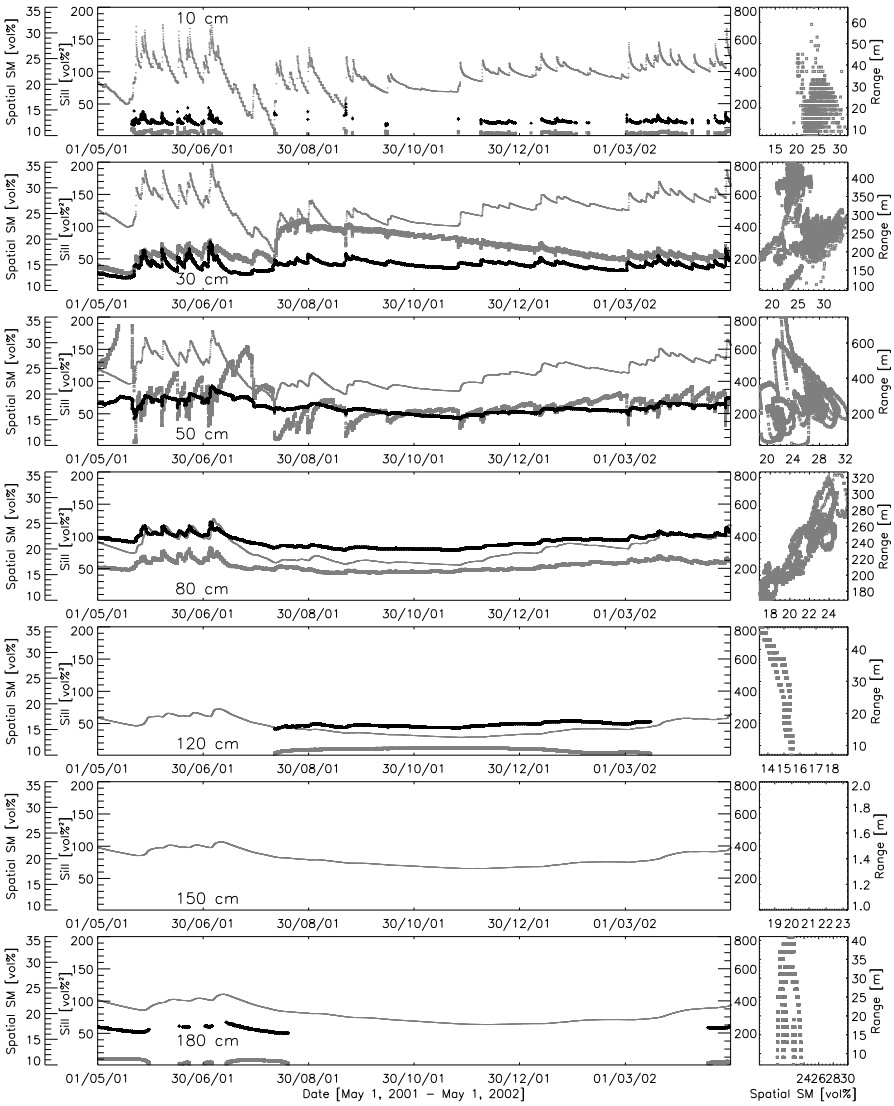


Figure 5.16: Time series of variogram parameters (range and sill in gray □ and black +, with right and left vertical scale, respectively) and spatially averaged soil moisture SM (thin gray, extra left vertical scale). The righthand side figures plot the range, observed at the different time steps, in function of the spatial average soil moisture at the corresponding time steps. Missing data occur when it was not possible to fit a variogram model.

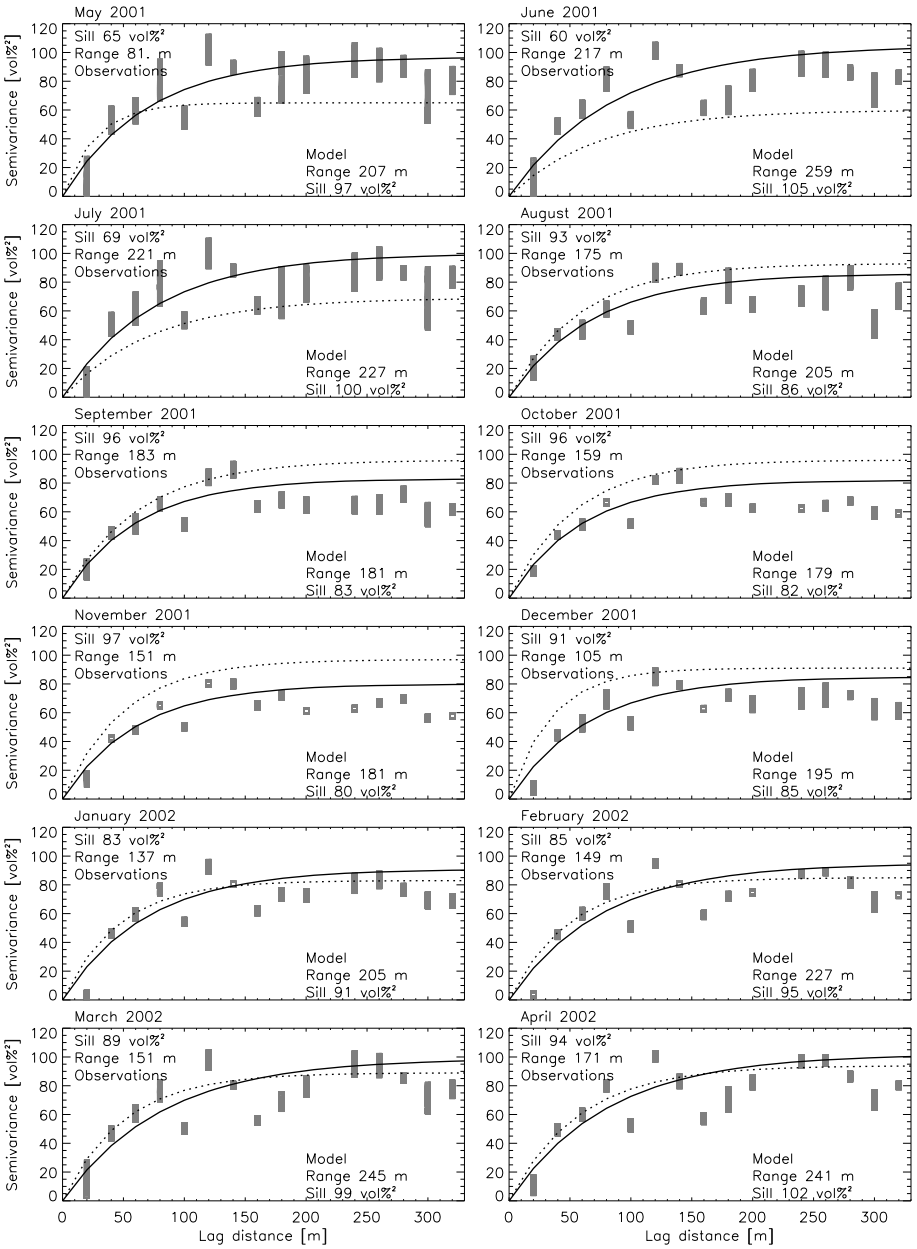


Figure 5.17: Pooled stratified (in time) variograms at 80 cm for each month in the period May 1, 2001 - April 30, 2002 for modeled soil moisture over the whole OPE^3 field (gray squares). The dashed line represents the best exponential variogram model fitted for observed soil moisture, while the full line was fitted for the modeled soil moisture.

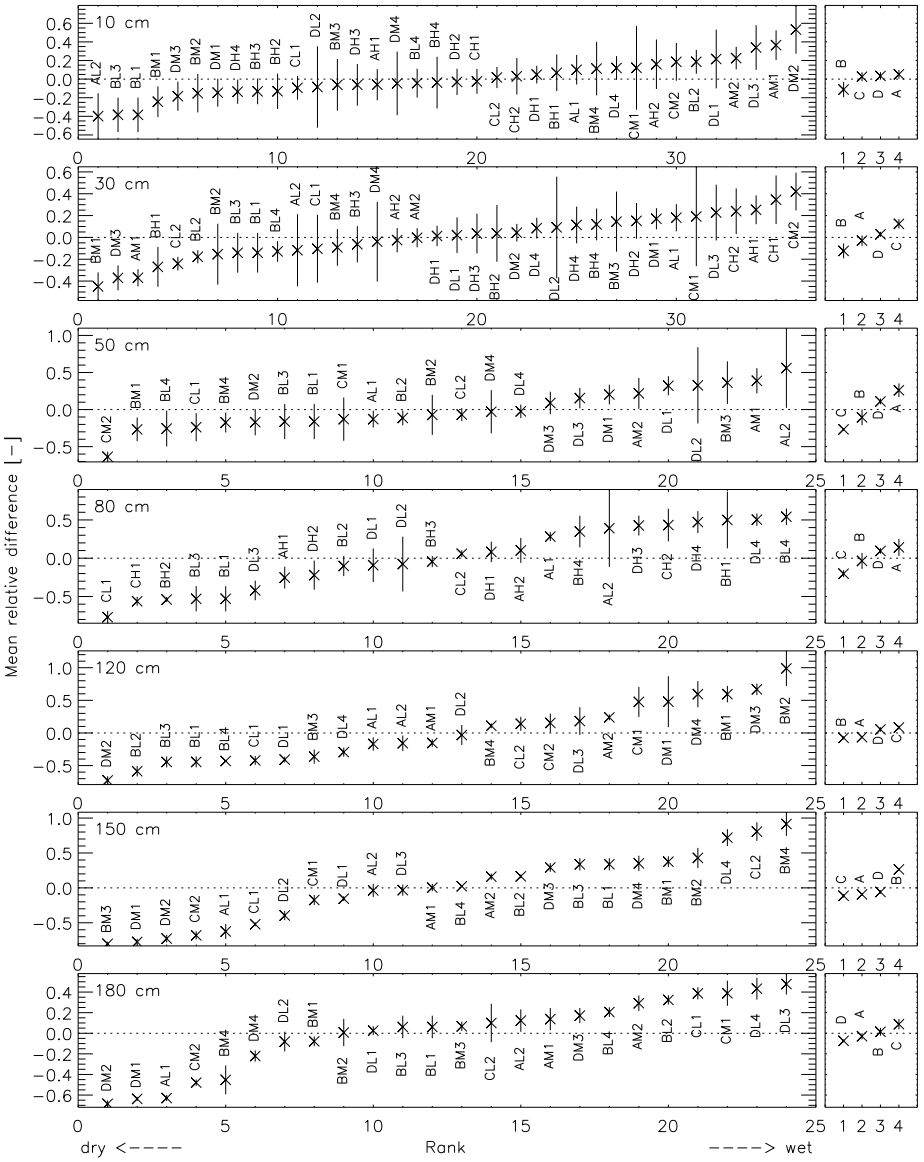


Figure 5.18: Average and standard deviation in relative differences from the areal wetness for all sensors (left) and for the individual fields (right) in the whole OPE^3 field. The sensors are sorted following their average deviation from the areal wetness. The average deviations are calculated by exact the same number of samples, i.e. 1 year (May 1, 2001 - April 30, 2002) of hourly data.

served spatial correlation by variogram fitting for the upper layer soil moisture, even though spatially constant input forcing was used. Apparently, the parameter assignment to the individual 1D profiles could not preserve the spatial structure of soil moisture in the field.

In the deeper layers, it was impossible to simulate the particular dry-out and wetting phenomena. In contrast to the upper layers, the main controls for the deeper layers were not known a priori and the simulations relied completely on the very simple model dynamics, which were insufficient to mimic complex flows. Also for the deepest layers, it was impossible to fit variogram models. The modeled variability in space was mostly less than observed in the deeper layers, which was exactly because the lateral flow was not modeled. Despite the inaccurate estimation of soil moisture in the deeper layers, it will be discussed in chapter 10 that the subsurface flow was relatively well simulated.

It is important to see that, for all soil layers, the model consistently generated higher spatial variability when the soil moisture was higher, while mainly for the deeper layers the opposite was observed. The modeled characteristic time scale was in general slightly larger than observed for all layers from 30 cm through 150 cm depth.

5.6 Distributed initialization and calibration

5.6.1 1D profiles

In the above discussion, the model was only calibrated for profiles where soil moisture data were available. If spatially averaged soil moisture would have been the required output, then it would have been best to run the model at individual points and to aggregate the output afterwards. Heuvelink and Pebesma [1999] explained that, due to the non-linearity of models, aggregating input before running a model does not yield the same result as running the model on individual points and aggregating the output afterwards. Heuvelink and Pebesma [1999] also studied how to deal with interpolation, separated from aggregation and discussed that there were two ways to obtain a (spatially) interpolated map of a variable. One could first interpolate the input and then run the model or run the model at individual sample sites and then interpolate the outputs. The authors found little difference between the 2 pathways, but stated that it is advised to first interpolate the input and then run the model, based on the fact that the spatial structure of the input is exploited best this way. For the study at hand, the input sensu stricto, i.e. the forcings, were assumed to be constant over the whole OPE^3 field. However, there is spatially variable input in broad sense through the parameterization (e.g. soil characteristics).

Modeling all points in the OPE^3 field could be achieved in 2 ways. A first option is to initially run and calibrate for the selected measurement points, and to in-

terpolate the parameters afterwards. The model can be run over all grid points, using interpolated parameters. This would be a realistic approach only if the found optimal parameters were really physically based. However, due to the unphysical nature of the parameters and their interactions, simple interpolation of the parameters is excluded. Furthermore, the combination of these interpolated parameters is even less likely to be optimal.

Therefore, a second and more viable option is to generate artificial soil moisture observations at all points in the field through interpolation and to calibrate for every single land column. Thus, in order to get a spatially representative control run, the model is to be calibrated for every single pixel. This way the model is calibrated for the imposed spatial variability, obtained through interpolation of the 'validating' observations.

To calibrate for all the grid cells after interpolation of the observations, has the main drawback of computational load. Further, it is quite possible that calibration of individual grid points causes spatial variability that is not natural, because it is not always possible to obtain an optimal fit with the observations and sometimes constant biases are found for some layers. In any case, the methods discussed here, consider the calibration of the individual profiles: it will be shown later that full 3D calibration is hardly (if not) feasible, when using a MC approach.

Since the CLM2.0 does not simulate horizontal flow, there is no need to simulate for the grid points in between the sensors: the model will not be able to propagate update information (obtained from data assimilation for state estimation) between grid cells (see chapter 8). Consequently, a simple interpolation of the model output as discussed in chapter 3 (a combination of the correlation pattern imposed for kriging and the pattern in texture and topindex related to soil moisture by a multivariate regression) was calculated where spatial views on the soil moisture were useful. Since the input (meteorological forcings) were assumed to be spatially uniform over the OPE^3 field, this approach was very reasonable. Spatially interpolated soil maps are shown in figure 5.19. The soil moisture patterns as obtained from interpolation of modeled soil moisture clearly differ from those obtained from observed data (figure 3.20), but the variability in time is similar.

5.6.2 2D/3D fields

The calibration was essentially performed in 1D in space, i.e. for soil profiles at individual points in space and by vertical aggregation of the objective functions over the different profile layers. If horizontal aggregation of the measures of goodness-of-fit would be performed over all sensors within the same layer, a calibration in 2D space could be performed. Such a spatial 2D calibration would probably better preserve the horizontal spatial structure. However, this option is not viable as per layer the optimal parameters would be found, but the combination of the parameters for the different layers will likely not represent a realistic profile, while the model is completely based on the calculation of the

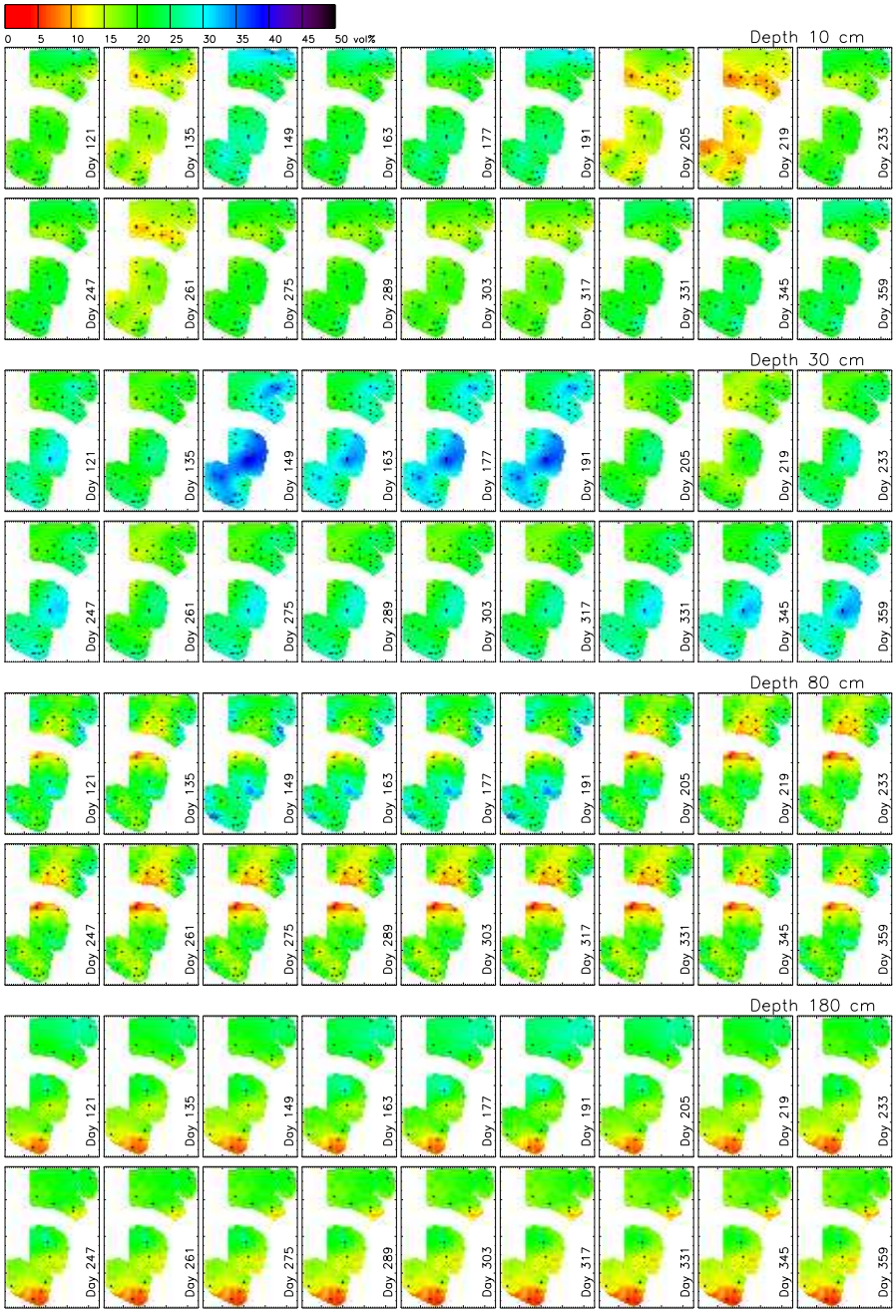


Figure 5.19: Spatially interpolated soil moisture at different layers in the OPE^3 field every 2 weeks, starting at May 1, 2001. Day numbers are counting from January 1, 2001. Black dots indicate where measurements were available for calibration.

profile soil moisture. Calibration in 3D would allow to account for the spatial structure of soil moisture in the horizontal plane and for the vertical distribution of water over the profile. It could for example be achieved by aggregating measures of goodness-of-fit in 3D and optimize them. However, the difficulty lays in the enormous amount of combinations of profiles parameters that should be considered for calibration. If again $15 \cdot 10^4$ MC simulations (patches or profiles) would be generated for the estimate of a single profile, this would result in $\frac{15 \cdot 10^4!}{(15 \cdot 10^4 - 36)!} = 15 \cdot 10^4 (15 \cdot 10^4 - 1) \ldots (15 \cdot 10^4 - 36 + 2)(15 \cdot 10^4 - 36 + 1)$ permutations (without repeated patches) to screen for the best parameter combination, if only 36 profiles were included. Even if the number of MC simulations would be reduced from $15 \cdot 10^4$ to any value larger than the amount of profiles, this amount of simulations would be excessive and therefore, full 3D calibration is not easily feasible, unless a priori information would be used, such as for example correlations between parameters.

5.7 Resulting statements

- The grid cell and patches structure of the CLM2.0 code was adapted and used to generate MC (global optimization) simulations for initialization and calibration, so that the parallel coding of the model could be optimally used.
- Parameter perturbations were chosen as Gaussian distributions around a physically realistic mean value and with a standard deviation of the magnitude of the value; additionally, some limitations are imposed as a priori information.
- Soil moisture observations from several soil layers and several different measures of goodness-of-fit were used in the parameter and initial state optimization; for each measure of goodness-of-fit, the objective functions for the different soil layers were combined into a single measure of goodness-of-fit for each profile.
- An iterative sorting procedure was used to select the best parameter vector based on a trade off between different measures of goodness-of-fit.
- Basically the combined initial state and parameter estimation was performed in a weak constraint variational assimilation scheme.
- For validation the same measures of goodness-of-fit were calculated as for the calibration, but for the period following the initialization (1 day) and calibration (1 month) period.
- Temporal and spatial characteristics as determined for observed soil moisture in chapter 3 were calculated for the modeled soil moisture; in general, temporal characteristics were found to be better preserved than spatial characteristics.

5.8 Summary

Through multi-objective initialization and calibration, a full identification of the CLM2.0 for the OPE^3 field was performed and a system describing the processes in the OPE^3 field as well as possible was obtained by assimilation of observed soil moisture data. There was however no indication that the final model was really the best one, given its structure. Many other parameters and initial conditions may also provide satisfactory results. Furthermore, through validation, it was shown that the model results showed several shortcomings and that for the OPE^3 field, soil moisture values were simulated that may be quite different in spatial structure.